Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ciencias Matemáticas
Unidad de Posgrado

Representación matricial de una aplicación lineal a través de la inversa generalizada de Moore-Penrose. Aplicaciones

TESIS

Para optar el Grado Académico de Magíster en Matemática Pura

AUTOR

Oswaldo Rafael LÓPEZ MICHELINI

ASESOR

Pedro Celso CONTRERAS CHAMORRO

Lima, Perú

2019

REPRESENTACIÓN MATRICIAL DE UNA APLICACIÓN LINEAL A TRAVÉS DE LA INVERSA GENERALIZADA DE MOORE-PENROSE. APLICACIONES

Oswaldo Rafael López Michelini

Tesis presentada a consideración del cuerpo de Docentes de la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos, como parte de los requisitos para obtener el Grado Académico de Magíster en Matemática Pura.

Aprobada por el jurado:

Dr. Alfonso Pérez Salvatierra
Presidente

Mg. Jorge Icaro Condado Jáuregui
Miembro

Mg. Josue Alonso Aguirre Enciso
Miembro

Mg. Luis Javier Vásquez Serpa
Miembro

Dr. Pedro Celso Contreras Chamorro
Miembro Asesor

Dedicatoria

El presente trabajo se lo dedico a Dios por ser el guía en todo mi camino y fortalecerme en los momentos mas difíciles, también a mi Madre Blanca Azuceti Michelini Mogollón y a mi hermano César López Michelini por el apoyo incondicional hacia mi persona.

Agradecimientos

Agradezco a Dios por darme la vida y la oportunidad de seguir realizando mis sueños.

Expreso mi más profundo agradecimiento al Dr. Pedro Celso Contreras Chamorro, por ser además de mi asesor, un gran amigo que siempre me mostró su ayuda, por haber aceptado dirigir este trabajo, por el invaluable apoyo brindado y especialmente por la paciencia a la hora de mi aprendizaje.

Al Dr. Alfonso Pérez Salvatierra, por haberme brindado la oportunidad de estudiar y profundizar todas aquellas materias que han sido de mi interés en estos dos años de estudios en la Unidad de Posgrado de la Escuela de Matemática.

Debo también agradecer de forma especial a mi madre, por su apoyo permanente y por el sacrificio que ha realizado para poder llevar a cabo mis estudios, mi formación académica y personal. Creo haber aprovechado cada uno de sus consejos y las oportunidades que con confianza me ha brindado, por todo ello le doy las gracias.

Doy también un sincero agradecimiento a mi hermano César López Michelini, por la disposición que te ha tenido para ayudarme en cada una de las adversidades que he pasado durante todo este tiempo.
Gracias hermano.

Mi gratitud también va dirigida a mi tía Elizabeth Yolanda Michelini Mogollón, en la misma medida debo un sincero reconocimiento a mi abuela Blanca Mogollón Infante por sus sabios consejos, por su cariño y por la confianza que ha tenido en mí en todo momento, a mi prima Blanca Michelini Cruz por el apoyo brindado a mi persona y a toda mi familia en general.

Por último, agradecer a la Mg. Roxana Rodríguez Escobedo por haberme introducido en este tema de investigación tan interesante, asimismo agradezco por la ayuda brindada al Dr. Wilson Maco Vásquez y al Dr. Franco Rubio López, todos ellos docentes de la Universidad Nacional de Trujillo, sin duda alguna puedo darles las gracias desde la admiración y la amistad.

Tabla de Símbolos

$\mathbb{K}$	:	Campo o cuerpo de los números reales $\mathbb{R}$ o complejos $\mathbb{C}$.
U, V, W	:	Espacios vectoriales sobre el campo $\mathbb{K}$.
V $\widetilde{\subset}$ W	:	W es subespacio de V.
N(T)	:	Núcleo del operador T.
R(T)	:	Rango del operador T.
$\oplus$	:	Suma directa.
$\mathcal{M}_{m\times n}(\mathbb{R})$	:	Espacio de matrices reales de orden m por n.
$A_{m\times n}(\mathbb{R})$	:	Indica que A es una matriz real de orden $m \times n$.
$\mathbb{R}^n$	:	Espacio euclideano n dimensional.
$\overline{\{1, n_0\}}$	:	Representación del conjunto $\{1, 2, 3, \cdots, n_0\}$
$Fil(A)$	:	Espacio fila asociado a la matriz A.
$Col(A)$	:	Espacio columna asociado a la matriz A.
$R_f(A)$	:	Rango fila de la matriz A.
$R_c(A)$	:	Rango columna de la matriz A.
$rg(A)$	:	Rango de la matriz A.
$A \sim B$	:	Equivalencia entre el objeto A y el objeto B.
$A^\dagger$	:	Inversa generalizada de Moore-Penrose de la matriz A.

$P_1 :\Leftrightarrow P_2$	:	Indica que P_1 equivale a P_2 por definición.
$P_1 \Leftrightarrow: P_2$	:	Indica que P_1 equivale a P_2 por notación.
$P_1 := P_2$	:	Indica que P_1 es igual a P_2 por definición.
$P_1 =: P_2$	:	Indica que P_1 es igual a P_2 por notación.
$\square$	:	Indica finalización de ejemplos.
$\blacksquare$	:	Indica finalización de demostraciones.

Contenido

Resumen

En el presente trabajo se ha construido la teoría formal de la inversa de Moore-Penrose, se ha determinado algunas expresiones de dicha inversa y se ha logrado dar una respuesta parcial a un problema que parece estar abierto.

Palabras clave: Suma directa, matriz proyección, inversa de Moore-Penrose, operador proyección.

Abstract

This paper depict the formal theory of Moore-Penrose inverse, some expressions have been determined and a parcial response do a problem that seems do be open.

Keywords: Direct sum, Projection matrix, Inverse of Moore Penrose, Projection operator.

Introducción

El concepto de **inversa generalizada** de una matriz lo introdujo por primera vez Eliakim Hastings Moore alrededor del año 1935, publicación que permaneció desconocida hasta la publicación del artículo [13] de Sir Roger Penrose en el año 1954.

Moore definió la inversa generalizada de una matriz, a través de proyecciones, haciendo así el trabajo de dicha matriz un tanto tediosa, por otro lado, Penrose definió la inversa generalizada de una matriz partiendo de cuatro ecuaciones matriciales, definición que hizo más sencilla la manipulación de dicha matriz. En 1955 R. Rado mostró que la definición de inversa generalizada emitida por Moore, es equivalente a la definición que dio Penrose. Es a partir de ese momento que dicha inversa generalizada comenzó a conocerse con el nombre de inversa de Moore-Penrose.[2]

La inversa de Moore-Penrose desempeña un papel fundamental en el análisis numérico moderno, puesto que su aplicación más notable es la de dar solución al problema de mínimos cuadrados.

El análisis numérico moderno está resolviendo sistemas de ecuaciones lineales usando la inversa de Moore-Penrose, pero sin tener un sustento teórico adecuado que permita construir con suficiente formalidad los algoritmos que están obteniéndose, en este sentido se está realizando un análisis numérico con una fundamentación matemática no formalizada, tal y como se hace en [19].

El concepto de la inversa de Moore-Penrose de una matriz, ha sido extendido a la teoría de operadores lineales acotados, es decir dado un operador lineal acotado definido entre dos espacios de Hilbert arbitrarios, es posible construir otro operador lineal acotado, llamado inverso de Moore-Penrose, el mismo que básicamente generaliza todas las propiedades del operador inverso, en caso de existir, tal y como se puede observar en [10].

Al realizar una lectura informal del inverso de Moore-Penrose en [10] y de la inversa de Moore-Penrose en [19], se concluye que ambos conceptos tienen el mismo comportamiento teórico, naturalmente, el primero a nivel de operadores lineales acotados y el segundo a nivel matricial.

Ahora, cuando se realiza el estudio del operador identidad y la matriz identidad, se observa que ambos objetos tienen el mismo comportamiento sobre sus respectivos espacios, y se puede mostrar que cuando el espacio es de dimensión finita, la matriz que representa al operador identidad justamente es la matriz identidad. Asimismo, cuando se realiza el estudio del operador nulo y la matriz nula se puede notar que sobre sus respectivos espacios ambos objetos tienen el mismo comportamiento, también sucede que, sobre espacios finito dimensionales la matriz asociada al operador nulo, precisamente es la matriz nula, esto induce a pensar que sobre espacios de dimensión finita se debe verificar que la matriz asociada al inverso de Moore-Penrose es la inversa de Moore-Penrose, esto es un problema que hasta el momento parece estar abierto.

Dar una respuesta parcial a dicho "problema abierto" y sustentar formalmente la forma de trabajo antes mencionada del análisis numérico moderno, lleva a la elaboración de los tres capítulos del presente trabajo, el cual tiene por objeto realizar un estudio profundo y formal de la inversa de Moore-Penrose.

En el capítulo 1, se incluyen los conceptos, definiciones y propiedades teóricas más relevantes de sobre matrices, operadores proyección, suma directa, etc.

En el capítulo 2 que es la parte central y más extensa de este trabajo, se inicia estudiando toda la teoría, que luego dentro de este mismo capítulo se utiliza para realizar la construcción formal de la inversa de Moore-Penrose, en esta misma parte del trabajo se demuestra detalladamente una serie de propiedades que tiene esta inversa, las mismas que llevan al teorema 2.34, resultado que da una respuesta parcial al problema abierto antes mencionado.

En el capítulo 3, se explica formalmente las aplicaciones que está tomando el análisis numérico moderno de la inversa de Moore-Penrose.

Oswaldo Rafael López Michelini

Lima, Febrero del 2019

Capítulo 1

Preliminares

En este capítulo se establecen algunos conceptos, notaciones, definiciones y resultados cuyo objetivo es fijar la base teórica que permite familiarizarse con la inversa generalizada de Moore−Penrose.

Cabe resaltar que los resultados que se desarrollan en esta parte no son estudiados a profundidad, ya que en su mayoría son estudiados en Álgebra Lineal. Sin embargo, es necesario repasarlos puesto que son la base de algunos de los capítulos siguientes.

Cabe mencionar que en el desarrollo de este trabajo, se hablará indistintamente de funciones y/o aplicaciones, es decir no se hará distinción entre estos conceptos, de la misma forma al hablar de operador deberá entenderse que se está haciendo referencia al concepto de operador lineal.

También es importante resaltar que el estudio a realizarse de esta parte, será peculiarmente sobre el campo de los números reales, debido a que facilita la comprensión de este nuevo tema de investigación.

Tener en cuenta que en todo el texto se usa como abreviatura de la palabra teorema, la sigla th.

1.1. Escalares y Espacio Euclidiano

Los **escalares** en $\mathbb{R}$, usualmente serán denotados con las letras griegas α, β, λ, etc.

Dado $n \in \mathbb{N}$, se define el conjunto $\mathbb{R}^n$ como el producto cartesiano de n factores iguales a $\mathbb{R}$, es decir:

$$\mathbb{R}^n := \underbrace{\mathbb{R} \times \mathbb{R} \cdots \times \mathbb{R}}_{n-veces}$$

Los puntos de $\mathbb{R}^n$ son n-uplas de la forma:

$$\overline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Cuyas **coordenadas** $x_1, x_2, \ldots, x_n$ son números reales.

Se define la *suma* en $\mathbb{R}^n$ de la siguiente manera:

$$\forall \overline{x}, \overline{y} \in \mathbb{R}^n : \overline{x} + \overline{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

Induciendo de este modo la siguiente función:

$$\begin{aligned} + : \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ (\overline{x}, \overline{y}) &\mapsto +(\overline{x}, \overline{y}) := \overline{x} + \overline{y} \end{aligned}$$

Asimismo, se define el *producto por un escalar* del siguiente modo:

$$(\forall \lambda \in \mathbb{R})(\forall \overline{x} \in \mathbb{R}^n) : \lambda \overline{x} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}$$

Induciendo también la función

$$\begin{aligned} \cdot : \mathbb{R} \times \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ (\lambda, \overline{x}) &\mapsto \cdot(\lambda, \overline{x}) := \lambda \cdot \overline{x} \end{aligned}$$

Como ya se sabe; el conjunto $\mathbb{R}^n$ dotado de estas dos operaciones $(\mathbb{R}^n, +, \cdot\mathbb{R}) =: \mathbb{R}^n$ es un *espacio vectorial de dimensión n sobre* $\mathbb{R}$.

Nota 1.1. *En el espacio vectorial $(\mathbb{R}^n, +, \cdot\mathbb{R}) =: \mathbb{R}^n$, se destaca la base estándar o también llamada base canónica, $\mathcal{E} := \{\overline{e}_1, \overline{e}_2, \ldots, \overline{e}_n\}$; donde:*

$$\forall j \in \overline{\{1, n\}} : \overline{e}_j := \begin{pmatrix} e_j(1) \\ e_j(2) \\ \vdots \\ e_j(n) \end{pmatrix}$$

siendo cada $e_j(k)$ el Delta de Kronecker; es decir $e_j(k) := \delta_{jk}$.

1.2. Producto Euclidiano y ortogonalidad

Producto Interno en $\mathbb{R}^n$

Definición 1.1. *Sean $\overline{x}$, $\overline{y}$ un par de vectores de $\mathbb{R}^n$. El producto interno de $\overline{x}$ con $\overline{y}$, está dado por:*

$$\langle \overline{x}, \overline{y} \rangle := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

Observación 1.1.

- ▷ *Como es ya de conocimiento usual, la definición anterior induce la siguiente función:*

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (\overline{x}, \overline{y}) &\mapsto \langle \overline{x}, \overline{y} \rangle \end{aligned}$$

  llamada Producto Euclidiano.

- ▷ *Se debe tener en cuenta las propiedades que esta aplicación satisface, ya que serán usadas en diversas situaciones.*

$\triangleright$ *En el presente trabajo se usará en forma continua la siguiente relación:*

$$\langle \overline{x}, \overline{y} \rangle =: \overline{x} \cdot \overline{y} := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Ortogonalidad en $\mathbb{R}^n$

Definición 1.2. *Sean $\overline{x}$, $\overline{y}$ un par de vectores de $\mathbb{R}^n$. El vector $\overline{x}$ es ortogonal al vector $\overline{y}$, $\overline{x} \perp \overline{y}$, si y sólo si $\langle \overline{x}, \overline{y} \rangle = 0$.*

Ortogonalidad: vector - conjunto

Definición 1.3. *Sea $\overline{x}$ un vector de $\mathbb{R}^n$ y A un subconjunto no vacío de $\mathbb{R}^n$. El vector $\overline{x}$ es ortogonal al conjunto A, $\overline{x} \perp A$, si y sólo si, para cada elemento $\overline{a} \in A$ se tiene que $\overline{x} \perp \overline{a}$.*

Ortogonalidad entre conjuntos

Definición 1.4. *Sean A, B un par de subconjuntos no vacíos de $\mathbb{R}^n$. El conjunto A es ortogonal al conjunto B, $A \perp B$, si y sólo si para cada $\overline{a} \in A$ y para cada $\overline{b} \in B$ se tiene que $\overline{a} \perp \overline{b}$.*

Sistema ortogonal

Definición 1.5. *Sea A un subconjunto no vacío de $\mathbb{R}^n$. El conjunto A es un sistema ortogonal si y sólo si para cada $\overline{x} \neq \overline{y} \in A$, se verifica que $\overline{x} \perp \overline{y}$.*

Nota 1.2. La expresión "sistema ortogonal" y "conjunto ortogonal" son sinónimas.

Teorema 1.1.
Si $\{\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_k\} \subset \mathbb{R}^n - \{\overline{0}\}$ es un conjunto ortogonal, entonces $\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_k$ son linealmente independientes.

Demostración: Revisar [9]

1.3. Complemento ortogonal y propiedades

Complemento ortogonal de un conjunto

Definición 1.6. *Sea A un subconjunto no vacío de $\mathbb{R}^n$. El complemento ortogonal de A, $A^{\perp}$, es el conjunto de vectores de $\mathbb{R}^n$ ortogonales a A.*

$$i.e. \quad A^{\perp} := \{\overline{x} \in \mathbb{R}^n / \, \overline{x} \perp A\}$$

Nota 1.3. *Considerando la definición anterior y la de ortogonalidad vector conjunto, se obtiene que:*

$$A^{\perp} := \{\overline{x} \in \mathbb{R}^n / \, \overline{x} \perp A\} := \{\overline{x} \in \mathbb{R}^n / \, \overline{x} \perp \overline{a}, \forall \overline{a} \in A\}$$

Teorema 1.2.
Sea W un subconjunto no vacío de $\mathbb{R}^n$. Si W es un subespacio, entonces:

1. *$W^{\perp}$ es un subespacio de $\mathbb{R}^n$.*
2. *$W \cap W^{\perp} = \{\overline{0}\}$.*
3. *$(W^{\perp})^{\perp} = W$.*
4. *$dim(W) + dim(W^{\perp}) = n$.*

Demostración: Revisar [14]

Nota 1.4. *Otra referencia en la que se puede revisar el resultado anterior es [8]*

1.4. Suma y suma directa

Suma de conjuntos en $\mathbb{R}^n$

Definición 1.7. *Sean A, B un par de subconjuntos no vacíos de $\mathbb{R}^n$. La suma de A y B está por:*

$$A + B := \{\overline{a} + \overline{b} / \, \overline{a} \in A \, , \, \overline{b} \in B\}$$

Nota 1.5. *Cuando A y B son subespacios de $\mathbb{R}^n$, se prueba que $A + B$ es también es un subespacio de $\mathbb{R}^n$.*

Suma directa en $\mathbb{R}^n$

Definición 1.8. *Sean V_1, V_2 un par de subespacios de $\mathbb{R}^n$. La suma de V_1 y V_2 es una suma directa , $V_1 \oplus V_2$,si y sólo si $V_1 \cap V_2 = \{\overline{0}\}$.*

$$i.e. \quad V_1 + V_2 = V_1 \oplus V_2 :\Leftrightarrow V_1 \cap V_2 = \{\overline{0}\}$$

Teorema 1.3.
Si W es un subespacio de $\mathbb{R}^n$ entonces $\mathbb{R}^n = W \oplus W^{\perp}$. Además, si M es un subespacio de $\mathbb{R}^n$ tal que $\mathbb{R}^n = W \oplus M$ y $M \perp W$ entonces $M = W^{\perp}$.

Demostración: Revisar [2]

Observación 1.2.
Usando los resultados anteriores se puede notar que, si:

$$V \tilde{\subset} \mathbb{R}^n \overset{th.1.2}{\Rightarrow} V^{\perp} \tilde{\subset} \mathbb{R}^n \overset{th.1.3}{\Rightarrow} \mathbb{R}^n = V^{\perp} \oplus (V^{\perp})^{\perp} \overset{th.1.2}{\Leftrightarrow} \mathbb{R}^n = V^{\perp} \oplus V$$

Esto quiere decir que:

$$\forall W \tilde{\subset} \mathbb{R}^n : \mathbb{R}^n = W \oplus W^{\perp} = W^{\perp} \oplus W.$$

Norma de un vector en $\mathbb{R}^n$

Definición 1.9. *Sea $\overline{x}$ un vector de $\mathbb{R}^n$. La norma de $\overline{x}$ está dada por:*

$$\|\overline{x}\| := \sqrt{\langle \overline{x}, \overline{x} \rangle} := \left(x_1^2 + x_2^2 + \cdots + x_n^2\right)^{\frac{1}{2}} =: \left(\sum_{j=1}^{n} x_j^2\right)^{\frac{1}{2}}$$

Observación 1.3.

- ▷ *Al igual que se ha visto en los conceptos anteriores, esta definición también induce la siguiente función.*

$$\begin{aligned} ||\cdot|| : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \overline{x} &\mapsto ||\cdot||(\overline{x}) := ||\overline{x}|| \end{aligned}$$

  llamada Norma Euclidiana.

- ▷ *En este trabajo continuamente se usara la siguiente relación: $\|\overline{x}\|^2 = \langle \overline{x}, \overline{x} \rangle$.*

- ▷ *Provisionando al espacio vectorial $\mathbb{R}^n$ del Producto Euclidiano y de la Norma Euclidiana, se obtiene el conocido Espacio Euclidiano, sobre el cual se trabajará.*

Teorema 1.4. *Teorema de Pitágoras*
Dados los vectores $\overline{a}$, $\overline{b}$ de $\mathbb{R}^n$ se verifica que:
1. $\overline{a} \perp \overline{b} \Leftrightarrow \left\|\overline{a}+\overline{b}\right\|^2 = \left\|\overline{a}\right\|^2 + \left\|\overline{b}\right\|^2$

2. $\overline{a} \perp \overline{b} \Leftrightarrow \left\|\overline{a}+\overline{b}\right\| = \left\|\overline{a}-\overline{b}\right\|$

Demostración: Revisar [12]

Conjunto ortonormal

Definición 1.10. *Sea A un subconjunto no vacío de $\mathbb{R}^n$. A es un conjunto ortonormal si y sólo si:*

$$\forall \overline{a}\,,\overline{b} \in A: \begin{cases} \overline{a} \perp \overline{b},\, \overline{a} \neq \overline{b} \\ ||\overline{a}|| = 1 \end{cases}$$

1.5. Proyección ortogonal y propiedades

Proyección ortogonal sobre un subespacio

Definición 1.11. *Sea W un subespacio de $\mathbb{R}^n$, $\overline{u}$ un vector fijo en $\mathbb{R}^n$ y $\overline{p}$ un vector en W. $\overline{p} \in W$ es la proyección ortogonal de $\overline{u}$ sobre W, $Proy_W(\overline{u})$, si y sólo si $\overline{u} - \overline{p} \perp W$.*

Teorema 1.5.
Sea W un subespacio de $\mathbb{R}^n$ y $\overline{u} \in \mathbb{R}^n$ un vector fijo. Si $\{\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_k\}$ es una base de W entonces:
1. $\exists!\, \overline{p} \in W / \overline{p} = Proy_W(\overline{u})$

2. Si $\{\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_k\}$ es una base ortogonal entonces:

$$\overline{p} = \frac{\langle \overline{u}_1, \overline{v}_1 \rangle}{||\overline{v}_1||^2} + \frac{\langle \overline{u}_2, \overline{v}_2 \rangle}{||\overline{v}_2||^2} + \cdots + \frac{\langle \overline{u}_n, \overline{v}_n \rangle}{||\overline{v}_n||^2}$$

3. Si la base $\{\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_k\}$ es ortonormal:

$$\overline{p} = \langle \overline{u}_1, \overline{v}_1 \rangle + \langle \overline{u}_2, \overline{v}_2 \rangle + \cdots + \langle \overline{u}_n, \overline{v}_n \rangle$$

Demostración: Revisar [4]

Nota 1.6. *Como se sabe que todo subespacio de $\mathbb{R}^n$ es finito dimensional, la importancia del resultado anterior para este trabajo está en el primer ítem, ya que está garantizando que la proyección ortogonal siempre existe y es única.*

Teorema 1.6. *Teorema de la mejor aproximación*
Sea W un subespacio de $\mathbb{R}^n$ y $\overline{y}_0 \in \mathbb{R}^n$ un vector fijo. Si $\widehat{y}$ es la proyección ortogonal de $\overline{y}_0$ sobre W entonces:

$$\forall \overline{w} \in W : ||\overline{y}_0 - \widehat{y}|| \leq ||\overline{y}_0 - \overline{w}||$$

Demostración: Revisar [4]

Observación 1.4. *El resultado anterior esta diciendo que el valor mínimo que toma la distancia $d(\overline{y}_0, \overline{w}) := ||\overline{y}_0 - \overline{w}||$ cuando $\overline{w}$ varía en W justamente se alcanza, cuando $\overline{v} = \widehat{y}$.* [4]

Teorema 1.7.
Si W un subespacio de $\mathbb{R}^n$ entonces:

1. $(\forall \overline{y} \in \mathbb{R}^n) \left(\exists!(\widehat{y}, z) \in W \times W^{\perp}\right) / \overline{y} = \widehat{y} + z$

2. $\widehat{y} = Proy_W(\overline{y})$

Demostración: Revisar [9]

1.6. Estudio matricial clásico

En el desarrollo del presente trabajo se ha considerado en forma usual que:

$$\mathcal{M}_{m\times n}(\mathbb{R}) := \left\{[a_{ij}] \Big/ i \in \overline{\{1, m\}}, j \in \overline{\{1, n\}}, a_{ij} \in \mathbb{R}\right\}$$

Se ha considerado también sobre dicho conjunto las siguientes definiciones.

$$\begin{cases} \forall [a_{ij}]_{m\times n}, [b_{ij}]_{m\times n} \in \mathcal{M}_{m\times n}(\mathbb{R}) : [a_{ij}]_{m\times n} + [b_{ij}]_{m\times n} := [a_{ij} + b_{ij}]_{m\times n} \\ (\forall \lambda \in \mathbb{R}) \left(\forall [a_{ij}]_{m\times n} \in \mathcal{M}_{m\times n}(\mathbb{R})\right) : \lambda [a_{ij}]_{m\times n} := [\lambda a_{ij}]_{m\times n} \end{cases}$$

Como ya se sabe, el conjunto $\mathcal{M}_{m\times n}(\mathbb{R})$ provisto de las operaciones anteriores es un espacio vectorial de dimensión finita sobre el campo $\mathbb{R}$. Y es el espacio sobre el cual este trabajo ha sido elaborado.

Observación 1.5.

▷ *Los elementos de $\mathcal{M}_{m\times n}(\mathbb{R})$ serán denotados como $A_{m\times n}(\mathbb{R})$, $B_{m\times n}(\mathbb{R})$, etc.*

▷ *El concepto de producto matricial utilizado en este trabajo, es el que se conoce en forma usual.*

▷ *La expresión A^T representará la transpuesta usual de la matriz $A_{m\times n}(\mathbb{R})$.*

1.6.1. Vectores asociados a una matriz

Si se supone que $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, entonces por definición se tiene que

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{1.1}$$

Luego, como para cada i, j se tiene que $a_{ij} \in \mathbb{R}$, entonces puede definirse los siguientes vectores de $\mathbb{R}^n$.

$$Col_1(A) := \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, Col_2(A) := \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \ldots, Col_n(A) := \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \tag{1.2}$$

De este modo, al usar (1.2) en (1.1), se puede ver que la matriz A quedará escrita de la siguiente manera.

$$A = \begin{bmatrix} Col_1(A) & Col_2(A) & \cdots & Col_n(A) \end{bmatrix} \tag{1.3}$$

Pensando en forma análoga, como para cada i, j se tiene que $a_{ij} \in \mathbb{R}$, entonces se puede definir las siguientes matrices.

$$\begin{cases} Fil_1(A) := \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{bmatrix} \\ Fil_2(A) := \begin{bmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{bmatrix} \\ \quad\vdots \\ Fil_m(A) := \begin{bmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \end{cases} \tag{1.4}$$

Por lo tanto, es claro que al usar (1.4) en (1.1) se tiene que la matriz A quedará escita del siguiente modo.

$$A = \begin{bmatrix} Fil_1(A) \\ Fil_2(A) \\ \vdots \\ Fil_n(A) \end{bmatrix} \tag{1.5}$$

Observación 1.6.

▷ *La expresión (1.5) es llamada representación vectorial fila de A y la expresión (1.3) es llamada representación vectorial columna de A.*

▷ *Para cada $j \in \overline{\{1,n\}}$, los vectores $Col_j(A)$ definidos en (1.2) son llamados vectores columna de A.*

▷ *Para cada $k \in \overline{\{1,m\}}$, los vectores $Fil_j(A)$ definidos en (1.4) son llamados vectores fila de A.*

1.6.2. Matriz ortogonal y propiedades

Matriz ortogonal

Definición 1.12. *Sea Q una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. Q es una matriz ortogonal si y sólo si $QQ^T = I_{n\times n} = Q^TQ$.*

Teorema 1.8.
Dada una matriz $Q_{n\times n}(\mathbb{R})$, se verifica que:

Q es una matriz ortogonal si y sólo si sus columnas son vectores ortonormales en $\mathbb{R}^n$.

Demostración: Revisar [14]

Teorema 1.9.
Dada una matriz $Q_{n\times n}(\mathbb{R})$. Los siguientes enunciados son equivalentes.

1. Q es una matriz ortogonal.

2. $\forall \overline{x} \in \mathbb{R}^n$: $||Q\overline{x}|| = ||\overline{x}||$

3. $\forall \overline{x}, \overline{y} \in \mathbb{R}^n$: $\langle Q\overline{x}, Q\overline{y}\rangle = \langle \overline{x}, \overline{y}\rangle$

Demostración: Revisar [14]

Teorema 1.10.
Dada una matriz $Q_{n\times n}(\mathbb{R})$. Si Q es una matriz ortogonal, entonces sus filas son vectores ortonormales en $\mathbb{R}^n$

Demostración: Revisar [14]

Teorema 1.11.
Sean Q, $\tilde{Q} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Si Q y $\tilde{Q}$ son matrices ortogonales, entonces $Q\tilde{Q}$ es una matriz ortogonal.

Demostración: Revisar [14]

Teorema 1.12.
Dada una matriz $Q_{m\times n}(\mathbb{R})$, se verifica que:
Las columnas de Q son vectores ortonormales en $\mathbb{R}^m$ si y sólo si $Q^T Q = I_{n\times n}$.

Demostración: Revisar [9]

1.6.3. Espacios asociados a una matriz

Espacio fila de una matriz

Definición 1.13. *Sea $A_{m\times n}(\mathbb{R})$ y sean $Fil_1(A)$, $Fil_2(A)$, $\ldots$, $Fil_m(A)$ los vectores fila de A. El espacio fila de A, es el espacio generado por dichos vectores fila.*

$$i.e. \quad Fil(A) := gen\,\{Fil_1(A), Fil_2(A), \ldots, Fil_m(A)\} \,\tilde{\subset}\, \mathbb{R}^n$$

Espacio columna de una matriz

Definición 1.14. *Sea $A_{m\times n}(\mathbb{R})$ y sean $Col_1(A)$, $Col_2(A)$, $\ldots$, $Col_n(A)$ los vectores columna de A. El espacio columna de A, es el espacio generado por dichos vectores columna.*

$$i.e. \quad Col(A) := gen\,\{Col_1(A), Col_2(A), \ldots, Col_n(A)\} \,\tilde{\subset}\, \mathbb{R}^m$$

Dimensiones de los espacios asociados a una matriz

Definición 1.15. *Sea* $A_{m\times n}(\mathbb{R})$*. Se llama*

- *rango fila de* A*,* $R_f(A)$*, a la dimensión del espacio fila de* A*.*

$$i.e. \quad R_f(A) := dim[Fil(A)]$$

- *rango columna de* A*,* $R_c(A)$*, a la dimensión del espacio fila de* A*.*

$$i.e. \quad R_c(A) := dim[Col(A)]$$

1.6.4. Rango de una matriz y propiedades

Teorema 1.13.
Si $A_{m\times n}(\mathbb{R})$ *entonces* $R_f(A) = R_c(A)$*, es decir* $dim[Fil(A)] = dim[Col(A)]$*.*

Demostración: Revisar [14]

Nota 1.7. *El teorema anterior muy importante en la teoría matricial, ya que él, permite hablar del rango de una matriz, sin tener que especificar si se trata del rango fila o rango columna, dando cabida a la siguiente definición.*

Rango de una matriz

Definición 1.16. *Sea* $A_{m\times n}(\mathbb{R})$*. El rango de* A*,* $rg(A)$*, está dado por el rango fila o el rango columna.*

$$rg(A) := R_f(A) = R_c(A)$$

o equivalentemente:

$$rg(A) := dim[Fil(A)] = dim[Col(A)]$$

Teorema 1.14.
Dada una matriz $A_{m\times n}(\mathbb{R})$ *y una matriz* $B_{n\times k}(\mathbb{R})$ *se cumple que:*

1. $rg(A) = rg(A^T)$*.*
2. *Si* $m = n$*,* A *es una matriz invertible si y sólo si* $rg(A) = n$*.*
3. $rg(AB) \leq min\{rg(A), rg(B)\}$*, es decir* $rg(AB) \leq rg(A)$ *y* $rg(AB) \leq rg(B)$*.*

Demostración: Revisar [11]

Observación 1.7. *Dada una matriz* $A_{m\times n}(\mathbb{R})$*, por definición se sabe que:*

- $Col(A) = gen\{Col_1(A), Col_2(A), \ldots, Col_n(A)\}$.
- $Fil(A) = gen\{Fil_1(A), Fil_2(A), \ldots, Fil_m(A)\}$.

Lo cual quiere decir que:

$$\begin{aligned} & dim[Col(A)] \leq n \quad \wedge \quad dim[Fil(A)] \leq m \\ :\Rightarrow\ & rg(A) \leq n \quad \wedge \quad rg(A) \leq m \\ \Rightarrow\ & rg(A) \leq min\{m, n\} \end{aligned}$$

1.7. Sistemas de ecuaciones y matrices

Se sabe que una ecuación lineal en las variables x_1, x_2, ..., x_n es una expresión de la forma:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \tag{1.6}$$

donde a_1, a_2, ... , a_n, $b \in \mathbb{R}$ por lo general son escalares conocidos, mientras que x_1, x_2, ... , $x_n \in \mathbb{R}$ en general son los escalares desconocidos.

Observación 1.8.

▷ *Los escalares conocidos* a_1, a_2, ... , a_n*, son llamados coeficientes de la ecuación lineal.*

▷ *El escalar* b*, es llamado término independiente.*

▷ *Los escalares* x_1, x_2, ... , x_n *son llamados incógnitas.*

1.7.1. Relación producto interno y producto matricial

Como en (1.6) se sabe que a_1, a_2, ... , a_n y x_1, x_2, ... , x_n son todos ellos números reales, por definición se tiene que existen los vectores $\overline{a}$, $\overline{b} \in \mathbb{R}^n$ de modo que:

$$\overline{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad , \quad \overline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

En la misma medida, se garantiza la existencia de la matriz:

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}^T := \overline{a}^{\,T}$$

Luego, usando la definición de producto interno y producto matricial en (1.6), se observa lo siguiente:

$$\begin{aligned} b &= a_1x_1 + a_2x_2 + \cdots + a_nx_n \\ &= \langle \overline{a}, \overline{x} \rangle \\ &:= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &:= \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &:= \overline{a}^{\,T} \cdot \overline{x} \end{aligned}$$

Observación 1.9. *Abstrayendo esta idea se tiene que:*

$$\forall \overline{x}\,,\, \overline{y} \in \mathbb{R}^n : \langle \overline{x}, \overline{y} \rangle = \overline{x}^{\,T} \cdot \overline{y}$$

Solución de una ecuación lineal

Definición 1.17. *Dada la ecuación lineal $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, un vector $\overline{s} \in \mathbb{R}^n$ es una solución de dicha ecuación lineal si y sólo si $\langle \overline{a}, \overline{s} \rangle = b$.*

$$i.e. \quad a_1s_1 + a_2s_2 + \cdots + a_ns_n = b$$

Nota 1.8. *A menudo, cuando $\overline{s} \in \mathbb{R}^n$ es una solución de la ecuación lineal $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, se escribe:*

$$x_1 = s_1,\, x_1 = s_1,\, \ldots\,,\, x_n = s_n$$

Conjunto solución de una ecuación lineal

Definición 1.18. *Dada la ecuación lineal* $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$*, el conjunto solución de dicha ecuación, es el conjunto formado por todos los* $\overline{s} \in \mathbb{R}^n$ *de modo que* $\langle \overline{a}, \overline{s} \rangle = b$.

$$i.e. \quad C.S := \{\overline{s} \in \mathbb{R}^n / \langle \overline{a}, \overline{s} \rangle = b\}$$

1.7.2. Análisis de la solución de una ecuación lineal

Dada la ecuación lineal:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \tag{1.7}$$

pueden presentarse los siguientes casos.

Primer caso

Si se supone que al menos uno de los coeficientes es no nulo, por ejemplo a_j, de (1.7) se tendría que

$$\begin{aligned}
&a_1x_1 + a_2x_2 + \cdots + a_{j-1}x_{j-1} + a_jx_j + a_{j+1}x_{j+1} + \cdots + a_nx_n = b \\
\Rightarrow\ &a_jx_j = b - a_1x_1 - a_2x_2 - \cdots - a_{j-1}x_{j-1} - a_{j+1}x_{j+1} - \cdots - a_nx_n \\
\Rightarrow\ &x_j = \frac{b}{a_j} - \frac{a_1}{a_j}x_1 - \frac{a_2}{a_j}x_2 - \cdots - \frac{a_{j-1}}{a_j}x_{j-1} - \frac{a_{j+1}}{a_j}x_{j+1} - \cdots - \frac{a_n}{a_j}x_n \\
\Rightarrow\ &\begin{cases}
x_1 : \text{arbitrario} \\
x_2 : \quad '' \\
\vdots \\
x_{j-1} : \quad '' \\
x_j = \dfrac{b}{a_j} - \dfrac{a_1}{a_j}x_1 - \dfrac{a_2}{a_j}x_2 - \cdots - \dfrac{a_{j-1}}{a_j}x_{j-1} - \dfrac{a_{j+1}}{a_j}x_{j+1} - \cdots - \dfrac{a_n}{a_j}x_n \\
x_{j+1} : \quad '' \\
\vdots \\
x_n : \text{arbitrario}
\end{cases}
\end{aligned}$$

Esto quiere decir que bajo esta condición, la ecuación lineal (1.7) tiene ∞^{n-1} soluciones.

Observación 1.10.

▷ *Debe entenderse que la notación* ∞^{n-1}, *está indicando que hay infinitas soluciones y que todas ellas se obtienen al hacer variar los* $(n-1)$ *parámetros arbitrarios* $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$.

▷ *Cabe recordar que en* (1.7), *cuando* $b = 0$ *se habla de una ecuación lineal homogénea, asimismo cuando* $b \neq 0$, *se habla de una ecuación lineal no homogénea.*

Segundo caso

Si se asume que todos los coeficientes son nulos, en (1.7) se tendrá que:

$$0x_1 + 0x_2 + \cdots + 0x_n = b$$

Ahora bien, si:

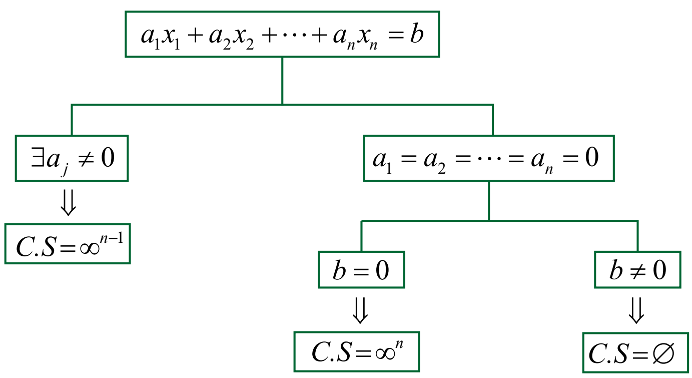
$$\begin{aligned} & b = 0 \\ \Rightarrow\ & 0x_1 + 0x_2 + \cdots + 0x_n = 0 \\ \Rightarrow\ & \begin{cases} x_1 : \text{arbitrario} \\ x_2 : \quad '' \\ \vdots \\ x_n : \quad '' \end{cases} \end{aligned}$$

Notando así que bajo esta hipótesis se tiene que existen ∞^n soluciones.

Por otro lado, si:

$$b \neq 0 \Rightarrow 0x_1 + 0x_2 + \cdots + 0x_n \neq 0$$

Concluyendo así que $C.S = \emptyset$.



Nota 1.9. *Todas estas ideas son generalizadas al estudiar los llamados sistemas de ecuaciones, llegando con dicho estudio al teorema de Rouché-Fröbenius.*

Sistema de ecuaciones lineales (S.E.L)

Definición 1.19. *Sean m, n un par de números naturales. Un sistema de "m" ecuaciones lineales en las variables x_1, x_2, $\ldots$, x_n de coeficientes a_{jk} con $j \in \overline{\{1, m\}}$, $k \in \overline{\{1, n\}}$ y términos independientes conocidos b_1, b_2, $\ldots$, b_m es aquel que se representa de la siguiente manera.*

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases} \tag{1.8}$$

Solución de un sistema de ecuaciones lineales

Definición 1.20. *Un vector $\overline{s} \in \mathbb{R}^n$ es una solución del sistema (1.8) si y sólo si dicho vector es solucion de cada una de las ecuaciones lineales que conforman el sistema.*

$$i.e. \quad \begin{cases} a_{11}s_1 & + & a_{12}s_2 & + & \cdots & + & a_{1n}s_n & = & b_1 \\ a_{21}s_1 & + & a_{22}s_2 & + & \cdots & + & a_{2n}s_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & \vdots \\ a_{m1}s_1 & + & a_{m2}s_2 & + & \cdots & + & a_{mn}s_n & = & b_m \end{cases}$$

Nota 1.10. *Al igual que para ecuaciones lineales, cuando $b_1 = b_2 = \cdots = b_m = 0$, se dice que (1.8) es un sistema lineal homogéneo, y cuando existe un $b_j \neq 0$, se dice que (1.8) es un sistema lineal no homogéneo.*

1.7.3. Representación matricial de un sistema de ecuaciones

Usando la definición de multiplicación matricial, se deduce en forma inmediata que (1.8) se puede expresar en forma equivalente de la siguiente manera.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{1.9}$$

Luego, al considerar que

$$A = [a_{ij}]_{m\times n} \quad , \quad \overline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \wedge \quad \overline{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

en (1.9) se tiene que:

$$\begin{cases} \quad A\,\overline{x} = \overline{b} \\ \text{S.a:} \\ \quad A \in \mathcal{M}_{m\times n}(\mathbb{R}) \\ \quad \overline{b} \in \mathbb{R}^m \\ \quad \overline{x} \in \mathbb{R}^n \text{ (Vector incógnita)} \end{cases} \tag{1.10}$$

Observación 1.11.

▷ *Se debe notar que (1.10), (1.9) y (1.8) representan exactamente el mismo problema.*

▷ *La matriz $A = [a_{ij}]_{m\times n}$ es llamada matriz de coeficientes.*

▷ *El vector $\overline{x}$ es llamado vector incógnita por ser el vector que se desconoce y $\overline{b}$ es llamado vector de términos independientes.*

Ahora bien, dado un sistema de ecuaciones lineales $A\overline{x} = \overline{b}$ es claro que:

- Un vector $\overline{s} \in \mathbb{R}^n$ es una solución de dicho sistema, si y sólo si $A\,\overline{s} = \overline{b}$.
- El sistema es homogéneo si y sólo si $\overline{b} = \overline{0} \in \mathbb{R}^m$.
- El conjunto solución, $C.S(A, \overline{b})$, está dado por $\left\{\overline{s} \in \mathbb{R}^n / A\,\overline{s} = \overline{b}\right\}$.

Clasificación de un S.E.L según su conjunto solución

Definición 1.21. *Dado un sistema de ecuaciones lineales $A\overline{x} = \overline{b}$, se dice que:*

1. *El sistema es incompatible, si y sólo si $C.S(A, \overline{b}) = \emptyset$.*
2. *El sistema es compatible, si y sólo si $C.S(A, \overline{b}) \neq \emptyset$, en este caso se dice que:*
   - *El sistema es compatible determinado, si y sólo si $\exists!\, \overline{s} \in C.S(A, \overline{b})$.*
   - *El sistema es compatible indeterminado, si y sólo si existen infinitas soluciones.*

Nota 1.11. *Dado un sistema lineal $A\overline{x} = \overline{b}$, el sistema $A\overline{x} = \overline{0}$ recibe el nombre de sistema homogéneo asociado.*

Teorema 1.15.
Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $\overline{b} \in \mathbb{R}^m$ y $A\overline{x} = \overline{b}$ un sistema de ecuaciones lineales con $\overline{x} \in \mathbb{R}^n$. Si $\overline{p} \in \mathbb{R}^n$ es una solución particular del sistema $A\overline{x} = \overline{b}$ entonces, el conjunto $M := \overline{p} + C.S(A, \overline{0})$ es el conjunto solución de dicho sistema.

$$i.e. \quad C.S(A, \overline{b}) = \overline{p} + C.S(A, \overline{0})$$

Demostración: Revisar [7] o [17]

Teorema 1.16.
Dada una matriz $A_{n\times n}(\mathbb{R})$ y un vector $\overline{b}$ fijo en $\mathbb{R}^n$, se cumple que:

1. *El sistema lineal $A\overline{x} = \overline{b}$ tiene solución única si y sólo si A es una matriz invertible. En este caso se verifica que $\overline{x} = A^{-1}\overline{b}$.*
2. *El sistema homogéneo $A\overline{x} = \overline{0}$ tiene solución no trivial si y sólo si , A es una matriz no invertible.*

Demostración: Revisar [18]

Teorema 1.17.
Dada una matriz $A_{m\times n}(\mathbb{R})$, se cumple que:

1. *Si $\overline{h}_1$, $\overline{h}_2$ son soluciones particulares del sistema homogéneo $A\overline{x} = \overline{0}$ y λ es un número real cualquiera, se tiene que:*
   - *$\overline{h}_1 + \overline{h}_2$ es una solución del sistema $A\overline{x} = \overline{0}$.*

- $\lambda\,\overline{h}_1$ *es una solución del sistema* $A\,\overline{x} = \overline{0}$.

2. *Si* $\overline{p}$ *es una solución particular del sistema no homogéneo* $A\,\overline{x} = \overline{b}$, *con* $\overline{b} \in \mathbb{R}^m$, *se tiene que:*
   - *Si* $\overline{h}$ *es una solución cualquiera del sistema homogéneo asociado* $A\,\overline{x} = \overline{0}$, *entonces* $\overline{p} + \overline{h}$ *es una solución del sistema no homogéneo* $A\,\overline{x} = \overline{b}$.
   - *Cualquier solución* $\overline{u}$ *del sistema no homogéneo* $A\,\overline{x} = \overline{b}$, *tiene la forma* $\overline{p}+\overline{h}$, *siendo* $\overline{h}$ *alguna solución del sistema homogéneo asociado* $A\,\overline{x} = \overline{0}$.

Demostración: Revisar [4]

Corolario 1.1.
Dada $A_{m\times n}(\mathbb{R})$ *se tiene que, el conjunto:*

$$Nul(A) := \{\overline{x} \in \mathbb{R}^n / A\,\overline{x} = 0\}$$

es un subespacio de $\mathbb{R}^n$.

Demostración: Revisar [12]

Teorema 1.18.
Sea $A_{m\times n}(\mathbb{R})$ *y sea* $\overline{b}$ *un vector cualquiera de* $\mathbb{R}^m$.

$$\overline{b} \in Col(A) \text{ si y sólo si el sistema } A\,\overline{x} = \overline{b} \text{ tiene alguna solución.}$$

Demostración: Revisar [4]

Corolario 1.2.
Dada $A_{m\times n}(\mathbb{R})$ *se tiene que:*

$$Col(A) = \left\{\overline{b} \in \mathbb{R}^m / \exists\,\overline{x} \in \mathbb{R}^n \,:\, A\,\overline{x} = \overline{b}\right\} = \{A\,\overline{x}/\overline{x} \in \mathbb{R}^n\}$$

Demostración: Revisar [12]

Observación 1.12.

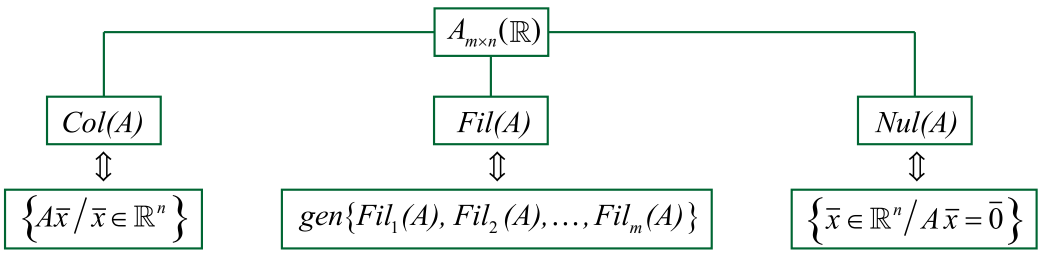
▷ *El corolario 1.1 se puede ver como una consecuencia directa del teorema 1.17 y el corolario 1.2 es consecuencia directa del teorema 1.18. Si dichos resultados no se desean ver como corolarios, se puede revisar las referencias indicadas.*

▷ *Cuando se tiene una función* $f : \mathbb{A} \to \mathbb{B}$, *se define la imagen de* f *como* $\mathrm{Im}(f) := \{f(x)/x \in \mathbb{A}\}$, *curiosamente este conjunto tiene mucha similitud a lo que se muestra en el corolario 1.2, pero esto no da derecho a hablar de la imagen de la matriz* A, *puesto que dicho término está reservado para funciones, y como puede verse en el corolario 1.2* A *es una matriz y no una función, por ello en este trabajo no se reconoce a dicho conjunto como la imagen de* A, *si no como el espacio columna de* A.

Espacio nulo de una matriz

Definición 1.22. *Sea* $A_{m\times n}(\mathbb{R})$. *El espacio nulo de* A, *está dado por el subespacio* $Nul(A) := \{\overline{x} \in \mathbb{R}^n / A\,\overline{x} = 0\}$.

Se puede resumir los espacios fundamentales asociados a una matriz en el siguiente esquema $A_{m\times n}(\mathbb{R})$.



Teorema 1.19. *Teorema del rango*

Dada una matriz $A_{m\times n}(\mathbb{R})$ *se verifica que:*

$$dim[Col(A)] + dim[Nul(A)] = n$$

Demostración: Revisar [12] o [14]

Teorema 1.20. *Teorema fundamental de matrices invertibles versión 2*

Dada una matriz $A_{n\times n}(\mathbb{R})$, *los siguientes enunciados son equivalentes.*

1. A *es una matriz invertible.*
2. $\forall\, \overline{b} \in \mathbb{R}^m$, *el sistema* $A\,\overline{x} = \overline{b}$ *tiene solución única.*
3. *El sistema* $A\,\overline{x} = \overline{0}$ *sólo tiene solución trivial.*
4. *La forma escalonada reducida de* A, *es la matriz identidad.*

5. *El rango de A es n. Es decir $rg(A) = n$.*
6. $dim[Nul(A)] = 0$.
7. *Los vectores columna de A son L.I.*
8. *Los vectores columna de A generan $\mathbb{R}^n$.*
9. *Los vectores columna de A forman una base de $\mathbb{R}^n$.*
10. *Los vectores fila de A son L.I.*
11. *Los vectores fila de A generan $\mathbb{R}^n$.*
12. *Los vectores fila de A forman una base de $\mathbb{R}^n$.*

Demostración: Revisar [14]

Nota 1.12. *Se recomienda revisar [14], ya que, en dicha referencia se encuentran un total de 5 versiones de este teorema, siendo la última de ellas la versión más completa.*

Teorema 1.21. *Teorema de Rouche-Fröbenius*
Sea $A_{m\times n}(\mathbb{R})$ y $\overline{b} \in \mathbb{R}^m$ un vector cualquiera. Si $A\,\overline{x} = \overline{b}$ es un sistema de ecuaciones lineales, se cumple que:

1. *El sistema es incompatible si y sólo si $rg(A) \neq rg(A|\overline{b})$.*
2. *El sistema es compatible si y sólo si $rg(A) = rg(A|\overline{b})$. En este caso se verifica que:*
   - *El sistema tiene solución única si y sólo si $rg(A) = \#$ incógnitas $= n$.*
   - *El sistema tiene infinitas soluciones, si y sólo si, se tiene que:* $rg(A) = rg(A|\overline{b}) < \#$ *incógnitas* $= n$.

Demostración: Revisar [6] o [1]

Nota 1.13. *En este trabajo se pretende complementar el resultado anterior, determinando quienes son las soluciones en caso de que existan.*

Teorema 1.22.

Dada una matriz $A_{m\times n}(\mathbb{R})$, se verifica que:

1. $Nul(A) \perp Col(A^T)$.
2. $Col(A^T) \oplus Nul(A) = \mathbb{R}^n$.
3. $Nul(A^T) \perp Col(A)$.
4. $[Col(A)]^{\perp} = Nul(A^T)$
5. $[Fil(A)]^{\perp} = Nul(A)$
6. $[Nul(A)]^{\perp} = Col(A^T)$
7. $Nul(A^T) = [Col(A)]^{\perp}$

Demostración: Revisar [2] y [14]

1.8. Autovalores y autovectores de una matriz

Autovalor de una matriz

Definición 1.23. *Sea $A_{n\times n}(\mathbb{R})$ y λ un número real cualquiera. El escalar λ es un autovalor de A si y sólo si, existe algún vector $\overline{x}$ no nulo en $\mathbb{R}^n$ de modo que $A\,\overline{x} = \lambda\,\overline{x}$.*

Observación 1.13.

▷ *Notar que la definición de valor propio sólo tiene sentido para matrices cuadradas.*

▷ *Si $\lambda \in \mathbb{R}$ es un valor propio de la matiz A y $\overline{u} \in \mathbb{R}^n - \{\overline{0}\}$ es un vector tal que $A\,\overline{u} = \lambda\,\overline{u}$, entonces se dice que $\overline{u}$ es un autovector de A, asociado λ.*

▷ *Tener en cuenta que la definición exige que el vector sea no nulo, pero del autovalor no exige nada.*

▷ *La matriz nula sólo tiene como autovalor al escalar cero.*

En efecto, dada la matriz nula $\Theta_{n\times n}(\mathbb{R})$, se tiene que, si λ es un autovalor de A, entonces por definicion se tiene que:

$$\begin{aligned}
&\exists\,\overline{x} \in \mathbb{R}^n - \{\overline{0}\} / \Theta\,\overline{x} = \lambda\,\overline{x} \\
:\Rightarrow\ & \lambda\,\overline{x} = \overline{0} \\
\overset{\text{th.}}{\Rightarrow}\ & \lambda = 0 \quad \vee \quad \underbrace{\overline{x} = \overline{0}}_{F} \\
\Rightarrow\ & \lambda = 0
\end{aligned}$$

▷ *Cada autovector, posee un único autovalor asociado.*

En efecto, dada una matriz $A_{n\times n}(\mathbb{R})$, se tiene que, si los escalares λ_1 y λ_2 son un par de autovalores distintos asociados al vector propio $\overline{x} \in \mathbb{R}^n - \{\overline{0}\}$, entonces por definición se tendría que:

$$
\begin{aligned}
& A\,\overline{x} = \lambda_1\,\overline{x} \quad \wedge \quad A\,\overline{x} = \lambda_2\,\overline{x} \\
\Rightarrow\; & \lambda_1\,\overline{x} = \lambda_2\,\overline{x} \\
\Rightarrow\; & (\lambda_1 - \lambda_2)\,\overline{x} = \overline{0} \\
\overset{\text{th.}}{\Rightarrow}\; & \lambda_1 - \lambda_2 = 0 \quad \vee \quad \underbrace{\overline{x} = \overline{0}}_{F} \\
\Rightarrow\; & \lambda_1 = \lambda_2 \quad (\Rightarrow\Leftarrow) \\
\therefore\; & \exists!\,\lambda_0 \in \mathbb{R} / A\,\overline{x} = \lambda_0\,\overline{x}
\end{aligned}
$$

▷ *Cada autovalor, posee infinitos autovectores asociados.*

En efecto, dada una matriz $A_{n\times n}(\mathbb{R})$, se tiene que, si los escalares λ_0 y es un autovalor de A y $\overline{x}$ es el autovector asociado a λ_0, por definición se tiene que:

$$
\begin{aligned}
A\,\overline{x} = \lambda_0\,\overline{x} &\Rightarrow \forall\, r \in \mathbb{R} - \{0\} : r(A\,\overline{x}) = r(\lambda_0\,\overline{x}) \\
&\Leftrightarrow \forall\, r \in \mathbb{R} - \{0\} : A\,\underline{(r\,\overline{x})} = \lambda_0\,\underline{(r\,\overline{x})}
\end{aligned}
$$

Esto quiere decir que para cada $r \in \mathbb{R} - \{0\}$, $r\,\overline{x}$ es un autovector asociado a λ_0. En consecuencia hay infinitos autovectores asociados a λ_0.

Nota 1.14.

1. *Las expresiones "valor propio", "valor característicos", "autovalor" y "eigenvalor" son sinónimas.*
2. *Las expresiones "vector propio", "vector característicos", "autovector" y "eigenvector" son sinónimas.*

Matrices equivalentes

Definición 1.24. *Sean A, B un par de matrices de orden $m \times n$ sobre $\mathbb{R}$. A es equivalente a B, $A \sim B$, si y sólo si, existe un par matrices invertibles P, Q de orden $m \times m$ y $n \times n$ respectivamente de modo que $A = P^{-1}BQ$.*

Teorema 1.23.
Sobre el espacio $\mathcal{M}_{m\times n}(\mathbb{R})$, se verifica que:

1. $\forall\, A \in \mathcal{M}_{m\times n}(\mathbb{R}) : A \sim A$.
2. $\forall\, A, B \in \mathcal{M}_{m\times n}(\mathbb{R}) : A \sim B \Rightarrow B \sim A$.
3. $\forall\, A, B, C \in \mathcal{M}_{m\times n}(\mathbb{R}) : A \sim B \,\wedge\, B \sim C \Rightarrow A \sim C$.

Demostración: Revisar [14]

Matrices semejantes

Definición 1.25. *Sean A, B un par de matrices cuadradas de orden $n \times n$ sobre $\mathbb{R}$. A es semejante a B, $A \sim B$, si y sólo si, existe alguna matriz invertible P de orden $n \times n$ de modo que $A = P^{-1}BP$.*

Teorema 1.24.
Sobre el espacio $\mathcal{M}_{n\times n}(\mathbb{R})$, se verifica que:

1. $\forall\, A \in \mathcal{M}_{n\times n}(\mathbb{R}) : A \sim A$.
2. $\forall\, A, B \in \mathcal{M}_{n\times n}(\mathbb{R}) : A \sim B \Rightarrow B \sim A$.
3. $\forall\, A, B, C \in \mathcal{M}_{n\times n}(\mathbb{R}) : A \sim B \,\wedge\, B \sim C \Rightarrow A \sim C$.

Demostración: Revisar [5]

Matrices congruentes

Definición 1.26. *Sean A, B un par de matrices cuadradas de orden $n \times n$ sobre $\mathbb{R}$. A es congruente a B, $A \sim B$, si y sólo si, existe alguna matriz invertible P de orden $n \times n$ de modo que $A = PBP^T$.*

Observación 1.14.

▷ *Puede probarse sin mayor dificultad que la congruencia de matrices define una relación de equivalencia sobre el espacio $\mathcal{M}_{n\times n}(\mathbb{R})$ al igual que las relaciones anteriores.*

▷ *Para cualquier par de matrices $A, B \in \mathcal{M}_{n\times n}(\mathbb{R})$, se cumple que, si A es semejante a B entonces $tr(A) = tr(B)$.*

En efecto, para esto hay que recordar que por teorema se sabe que:

$$(\forall X \in \mathcal{M}_{m\times n}(\mathbb{R}))(\forall Y \in \mathcal{M}_{n\times m}(\mathbb{R})) : tr(XY) = tr(YX)$$

Luego, si:

$$\begin{aligned} A \sim B &\Rightarrow \exists\, P \in \mathcal{M}_{n\times n}(\mathbb{R}) / A = P^{-1}BP \\ &\Rightarrow tr(A) = tr(\underbrace{P^{-1}}_{X}\underbrace{BP}_{Y}) = tr(BPP^{-1}) = tr(B) \\ &\Rightarrow tr(A) = tr(B) \end{aligned}$$

▷ *Si $A, B \in \mathcal{M}_{n\times n}(\mathbb{R})$, son congruentes, entonces son equivalentes.*

En efecto:

$$\begin{aligned} A \sim B &\Rightarrow \exists P \in \mathcal{M}_{n\times n}(\mathbb{R}) / A = PBP^T \\ &\Rightarrow B = P^{-1}A(P^T)^{-1} \\ &\Rightarrow B = P^{-1}A(P^{-1})^T \underset{\underset{(P^{-1})^T = Q}{\uparrow}}{=} P^{-1}AQ \\ &\Rightarrow B = P^{-1}A\,Q \end{aligned}$$

▷ *Si $A, B \in \mathcal{M}_{n\times n}(\mathbb{R})$, son semejantes, entonces son equivalentes.*

En efecto:

$$A \sim B \Rightarrow \exists P \in \mathcal{M}_{n\times n}(\mathbb{R}) / A = P^{-1}AP \underset{\underset{Q=P}{\uparrow}}{\Rightarrow} A = P^{-1}A\,Q$$

1.9. Diagonalización de matrices y propiedades

Matriz diagonalizable

Definición 1.27. *Sea $A_{n\times n}(\mathbb{R})$. A es una matriz diagonalizable si y sólo si es semejante a una matriz diagonal, es decir siempre y cuando exista alguna matriz invertible $P_{n\times n}(\mathbb{R})$, de modo que $D = P^{-1}A\,P$ es una matriz diagonal.*

Observación 1.15.

▷ *Cualquier matriz diagonal $D := diag(d_{11}, d_{22}, \ldots, d_{nn})$ es diagonalizable, pues $D = I_{n\times n}^{-1}\, D\, I_{n\times n}$.*

▷ *Cuando una matriz $A_{n\times n}(\mathbb{R})$ es diagonalizable, se suele decir que A puede diagonalizarse.*

▷ *La matriz P de la definición anterior, suele llamarse matriz diagonalizante.*

Teorema 1.25.

Dada una matriz $A_{n\times n}(\mathbb{R})$, se verifica que:

$$A \text{ es invertible si y sólo si } 0 \in \mathbb{R} \text{ no es un autovalor de } A.$$

Demostración: Revisar [14]

Teorema 1.26.

Sea $A_{n\times n}(\mathbb{R})$. Si λ es un autovalor de A y $\overline{x}_0 \in \mathbb{R}^n - \{\overline{0}\}$ es el autovector asociado a λ entonces:

1. *Para cualquier $m \in \mathbb{N}$, se tiene que λ^m es un autovalor de A^m, cuyo autovector asociado es $\overline{x}_0$.*
2. *Si A es una matriz invertible, entonces $\dfrac{1}{\lambda}$ es un autovalor de A^{-1}, cuyo autovector asociado es $\overline{x}_0$.*
3. *Si A es una matriz invertible, entonces para cualquier $m \in \mathbb{Z}$, se tiene que λ^m es un autovalor de A^m, cuyo autovector asociado es $\overline{x}_0$.*

Demostración: Revisar [14]

1.9.1. Ecuación característica y polinomio característico

Polinomio característico / ecuación característica

Definición 1.28. *Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$.*

1. *El polinomio característico de A, $P_A(\lambda)$, está dado por el determinante de la matriz $A - \lambda I$, i.e. $P_A(\lambda) := det(A - \lambda I)$.*
2. *La expresión $P_A(\lambda) = 0$, recibe el nombre de ecuación característica.*

Observación 1.16.

- ▷ *Las expresiones “ecuación característica”, “autoecuación”, “ecuación propia” y “eigenecuación” son sinónimas.*
- ▷ *Las expresiones “polinomio característico”, “autopolinomio”, “polinomio propio” y “eigenpolinomio” son sinónimas.*
- ▷ *La matriz $A - \lambda I$, suele conocerse con el nombre de matriz característica de A.*

Teorema 1.27.
Si A es una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$, entonces:

$$P_A(\lambda) = (-1)^n \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$$

donde para cada $j \in \overline{\{1, n\}}$: $c_j \in \mathbb{R}$, $P_A(\lambda)$ es un polinomio de grado n en la variable λ y además $c_{n-1} = (-1)^n \, tr(A)$.

Demostración: Revisar [4]

Teorema 1.28.
Si $A_{n\times n}(\mathbb{R})$, entonces los autovalores de A son las raíces de su polinomio característico $P_A(\lambda)$.

Demostración: Revisar [15]

Teorema 1.29.
Dada una matriz $A_{n\times n}(\mathbb{R})$ se tiene que, $\lambda_0 \in \mathbb{R}$ es un autovalor de A si y sólo si λ_0 es un autovalor de A^T.

Demostración: Revisar [12]

Teorema 1.30.
Sea $A_{n\times n}(\mathbb{R})$, α, β escalares distintos de $\mathbb{R}$ y sean $\overline{x}$, $\overline{y}$ un par de vectores no nulos de $\mathbb{R}^n$.
Si α es un autovalor de A con autovector asociado $\overline{x}$ (i.e. $A\,\overline{x} = \alpha\,\overline{x}$) y β es autovalor de A^T con autovector asociado $\overline{y}$ (i.e. $A^T\,\overline{y} = \beta\,\overline{y}$), entonces $\overline{x} \perp \overline{y}$.

Demostración: Revisar [12]

Espectro de una matriz

Definición 1.29. *Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. El espectro de A , $sp(A)$, está dado por el conjunto de todos sus autovalores reales o complejos.*

Multiplicidad algebraica de un autovalor

Definición 1.30. *Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$, $\lambda_0 \in \mathbb{R}$ un autovalor de A y sea m_0 un número natural. λ_0 es un autovalor de multiplicidad algebraica m_0, $ma(\lambda_0) = m_0$, si y sólo si, $P_A(\lambda) = (\lambda - \lambda_0)^{m_0} q(\lambda)$, siendo $q(\lambda)$ un polinomio que no tiene a λ_0 como raíz.*

Nota 1.15. *En lugar de hablar de la multiplicidad algebraica de λ_0, simplemente se habla de la multiplicidad de λ_0.*

Teorema 1.31.
Dada una matriz $A_{n\times n}(\mathbb{R})$ se verifica que:

1. *Si $\lambda_0 \in \mathbb{R}$ es un autovalor de A y $E(\lambda_0) := \left\{\overline{x} \in \mathbb{R}^n - \{\overline{0}\} / A\,\overline{x} = \lambda_0\,\overline{x}\right\} \cup \{\overline{0}\}$, entonces $(E(\lambda_0), +, \cdot\mathbb{R}) =: E(\lambda_0)$ es un subespacio de $\mathbb{R}^n$.*
2. *Si $\lambda_0, \lambda_1 \in \mathbb{R}$ son autovalores distintos de A, entonces $E(\lambda_0) \cap E(\lambda_1) = \{\overline{0}\}$.*
$$i.e. \quad \lambda_0 \neq \lambda_1 \Rightarrow E(\lambda_0) \cap E(\lambda_1) = \left\{\overline{0}\right\}$$
3. *Autovectores asociados a autovalores diferentes son linealmente independientes.*

Demostración: Revisar [3]

Observación 1.17.

▷ *Del teorema anterior se deduce que:*

$$\begin{aligned} E(\lambda_0) &:= \left\{\overline{x} \in \mathbb{R}^n - \{\overline{0}\} / A\,\overline{x} = \lambda_0\,\overline{x}\right\} \cup \{\overline{0}\} \\ &= \{\overline{x} \in \mathbb{R}^n / A\,\overline{x} = \lambda_0\,\overline{x}\} \\ &= \left\{\overline{x} \in \mathbb{R}^n / A\,\overline{x} - \lambda_0\,\overline{x} = \overline{0}\right\} \\ &= \left\{\overline{x} \in \mathbb{R}^n / (A - \lambda_0\,I)\,\overline{x} = \overline{0}\right\} \\ &:= Nul(A - \lambda_0\,I) \end{aligned}$$

▷ *El espacio $E(\lambda_0)$ algunas ocaciones es denotado por $V_{\lambda_0}(A)$ o $E_{\lambda_0}(A)$.*

Autoespacio / Subespacio propio

Definición 1.31. *Sea A una matriz cuadrada de orden $n\times n$ sobre $\mathbb{R}$, y $\lambda_0 \in \mathbb{R}$ un autovalor de A. El subespacio propio de A, asociado a λ_0, está dado por el subespacio de $\mathbb{R}^n$, $(E(\lambda_0), +, \cdot\mathbb{R}) =: E(\lambda_0)$.*

Multiplicidad geométrica

Definición 1.32. *Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$, y $\lambda_0 \in \mathbb{R}$ un autovalor de A. La multiplicidad geométrica de λ_0, $mg(\lambda_0)$, está dada por la dimensión del subespacio $(E(\lambda_0), +, \cdot\mathbb{R}) =: E(\lambda_0)$, i.e. $mg(\lambda_0) := dim[E(\lambda_0)]$.*

Teorema 1.32.
Sea $A_{n\times n}(\mathbb{R})$. Si $\lambda_0 \in \mathbb{R}$ un autovalor de A entonces se tiene que:

$$1 \leq mg(\lambda_0) := dim[E(\lambda_0)] \leq ma(\lambda_0)$$

Demostración: Revisar [14]

1.9.2. Criterios de diagonalización

Teorema 1.33.
Dada una matriz cuadrada A de orden $n \times n$ sobre, se cumple que:

1. *Si A tiene n autovectores $\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_n \in \mathbb{R}^n$ linealmente independientes asociados a los autovalores $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectivamente, entonces existen las matrices invertibles $P = \begin{pmatrix} \overline{v}_1 & \overline{v}_2 & \cdots & \overline{v}_n \end{pmatrix}$ y $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ de modo que $A = P\,D\,P^{-1}$, es decir A es diagonalizable.*

2. *Recíprocamente, si A es diaginalizable entonces A tiene n autovectores linealmente independientes en $\mathbb{R}^n$.*

Demostración: Revisar [18]

Teorema 1.34.
Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$.
Si $\overline{v}_1(1), \overline{v}_2(1), \ldots, \overline{v}_{n1}(1)$ son autovectores linealmente independientes asociados a λ_1, $\overline{v}_1(2), \overline{v}_2(2), \ldots, \overline{v}_{n2}(2)$ son autovectores linealmente independientes asociados a λ_2, $\ldots$, $\overline{v}_1(k), \overline{v}_2(k), \ldots, \overline{v}_{nk}(k)$ son autovectores linealmente independientes asociados a λ_k y $\lambda_1, \lambda_2, \ldots, \lambda_k$ son autovalores distintos de A entonces el conjunto $\{\overline{v}_1(1), \overline{v}_2(1), \ldots, \overline{v}_{n1}(1), \overline{v}_2(2), \ldots, \overline{v}_{n2}(2), \ldots, \overline{v}_1(k), \overline{v}_2(k), \ldots, \overline{v}_{nk}(k)\}$ es L.I.

Demostración: Revisar [18]

Teorema 1.35.
Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. Si A tiene n autovalores distintos entonces A es diagonalizable.

Demostración: Revisar [6]

Teorema 1.36.
Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. A es diagonalizable, si y sólo si $\mathbb{R}^n$ tiene una base autovectores de A.
Además en la descomposición $D = P^{-1}\,A\,P$, las columnas de la matriz P son los autovectores de A y los elementos diagonales de D, son los autovalores de A.

Demostración: Revisar [11] o [4]

Teorema 1.37. *Teorema de diagonalización*
Sea A una matriz de $n \times n$ sobre $\mathbb{R}$. Si $\lambda_1, \lambda_2, \ldots, \lambda_k$ son autovalores distintos de A entonces los siguientes enunciados son equivalentes.

1. *A es diagonalizable.*
2. *La unión B de las bases de los autoespacios de $\lambda_1, \lambda_2, \ldots, \lambda_k$, contiene n autovectores.*
3. $\forall\, j \in \overline{\{1,k\}} : ma(\lambda_j) = mg(\lambda_j) := dim[E_{\lambda_j}(A)]$.

Demostración: Revisar [14]

Matriz simétrica

Definición 1.33. *Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. A es una matriz simétrica, si y sólo si $A = A^T$.*

Matriz diagonalizable ortogonalmente

Definición 1.34. *Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. A es es diagonalizable ortogonalmente, si y sólo si existe alguna matriz ortogonal $Q \in \mathcal{M}_{n\times n}(\mathbb{R})$, de modo $D := Q^T\,A\,Q$ es una matriz diagonal.*

Nota 1.16. *La definición anterior está diciendo que A es diagonalizable por medio de una matriz ortogonal.*

Teorema 1.38.
Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. Si A es una matriz diagonalizable ortogonalmente, entonces A es una matriz simétrica.

Demostración: Revisar [14]

Teorema 1.39. *Teorema espectral*
Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. A es simétrica, si y sólo si, A es ortogonalmente diagonalizable.

Demostración: Revisar [14]

Teorema 1.40.
Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. Si A es una matriz simétrica, entonces existe alguna matriz ortogonal $Q \in \mathcal{M}_{n\times n}(\mathbb{R})$, de modo que $D := Q^T A\, Q$ es una matriz diagonal formada por los autovalores de A, algunos de los cuales pueden repetirse.

Demostración: Revisar [5]

Teorema 1.41.
Dada una matriz $A_{n\times n}(\mathbb{R})$, se verifica que:

1. *Si A es una matriz simétrica, entonces A es diagonalizable.*
2. *Si A es una matriz simétrica e idempotente $(A^2 = A)$, entonces $rg(A) = tr(A)$.*

Demostración: Revisar [12]

Teorema 1.42.
Dada una matriz $A_{n\times n}(\mathbb{R})$, se verifica que:

1. *Si A es diagonalizable entonces A tiene n autovalores no necesariamente diferentes.*
2. *A es diagonalizable si y sólo si existe una base de $\mathbb{R}^n$ formada por autovectores de A.*

Demostración: Revisar [3]

Teorema 1.43. *Teorema espectral para matrices simétricas*
Sea A una matriz cuadrada de orden $n \times n$ sobre $\mathbb{R}$. Si A es simétrica entonces $\mathbb{R}^n$ tiene una base ortonormal formada por autovectores de A.

Demostración: Revisar [11]

1.10. Estudio matricial generalizado

1.10.1. Matrices por bloques / Matrices particionadas

La manipulación de matrices con un gran número de filas y columnas conlleva grandes problemas, incluso cuando son manejados en un ordenador. Por esta razon es importante poder descomponer problemas que usan matrices "grandes" en otros problemas que utilizan matrices mucho más "pequeñas".

Nota 1.17. *La teoría relacionada a las matrices por bloques o matrices particionadas puede hallarse en [16] y [12].*

Submatriz: Eliminación de filas y columnas

Definición 1.35. *Sean A, A_0 un par de matrices reales de orden $m \times n$ y $m' \times n'$ respectivamente con $m' < m$ y $n' < n$. A_0 es una submatriz de A, $A_0 \rightarrowtail A$, si y sólo si A_0 se obtiene eliminando r-filas y s-columnas de A.*

Nota 1.18. *Para indicar que A_0 se obtuvo eliminando la fila 1, 2, $\dots$, i, $\dots$, r y la columna 1, 2, $\dots$, j, $\dots$, s de A, se usa la siguiente expresión.*

$$A(1, 2, \dots, i, \dots, r|1, 2, \dots, j, \dots, s)$$

donde i representa la posición de la fila a eliminar y j representa la posición de la columna a eliminar.

Ejemplo 1.1. *Dada la matriz:*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix}$$

determine la submatriz $A(2,4|2,4,5)$.

Solución:

La expresión $A(2,4|2,4,5)$ está indicando que debe eliminarse $Fil_2(A)$, $Fil_4(A)$, $Col_2(A)$, $Col_4(A)$ y $Col_5(A)$.

Al realizar el proceso anterior por medio de lineas verticales y horizontales sobre la matriz A, se tiene que:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix}$$

Por lo tanto $A(2,4|2,4,5) = \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix}$

□

Submatriz: Intersección de filas y columnas

Definición 1.36. *Sean A, A_0 un par de matrices reales de orden $m \times n$ y $m' \times n'$ respectivamente con $m' \leq m$ y $n' \leq n$. A_0 es una submatriz de A, $A_0 \rightarrowtail A$, si y sólo si A_0 se obtiene interceptando r-filas y s-columnas de A.*

Nota 1.19. *Para indicar que A_0 se obtuvo intercectando la fila 1, 2, . . . , i, . . . , r y la columna 1, 2, . . . , j, . . . , s de A, se usa la siguiente expresión.*

$$A[1,2,\ldots,i,\ldots,r|1,2,\ldots,j,\ldots,s]$$

donde i representa la posición de la fila a intersectar y j representa la posición de la columna a intersectar.

Ejemplo 1.2. *Dada la matriz:*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$$

determine la submatriz $A[1,2|3]$.

Solución:

La expresión $A[1,2|3]$ está indicando que debe intersectarse $Fil_1(A)$ y $Fil_2(A)$ con $Col_3(A)$.

Al realizar el proceso anterior por medio de lineas verticales y horizontales sobre la matriz A, se tiene que:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$$

Por lo tanto $A[1,2|3] = \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix}$

$\square$

1.10.2. Partición de una matriz

Sea $A_{m\times n}(\mathbb{R})$, para realizar una partición sobre A, puede seguirse el siguiente proceso.

1. Hacer $m = m_1 + m_2 + \cdots + m_r$ de modo que:

m_1 : Indique que la primer agrupación consta de las m_1 primeras filas.

m_2 : Indique que la segunda agrupación consta de las m_2 siguientes filas.

$\vdots$

m_r : Indique que la última agrupación consta de las m_r últimas filas.

2. Hacer $n = n_1 + n_2 + \cdots + n_s$ de modo que:

n_1 : Indique que la primer agrupación consta de las n_1 primeras columnas.

n_2 : Indique que la segunda agrupación consta de las n_2 siguientes columnas.

$\vdots$

n_s : Indique que la última agrupación consta de las n_s últimas columnas.

Nota 1.20. *Las descomposiciones $m = m_1+m_2+\cdots+m_r$, $n = n_1+n_2+\cdots+n_s$ son llamadas patrones de partición.*

Ejemplo 1.3. *Dada la matriz:*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ a_{51} & a_{52} & a_{53} & a_{54} \end{bmatrix}$$

determine la partición de ella, según el patrón de partición:

$$\begin{cases} 5 = m = 3 + 2 \\ 4 = n = 1 + 2 + 1 \end{cases}$$

Solución:

La expresión $m = 3 + 2$ indica que en la partición, la primer agrupación debe constar de 3 filas y la última agrupación debe contener 2 filas, asimismo, la expresión $n = 1 + 2 + 1$ señala que en la partición la primer agrupación debe contener 1 columna, la segunda agrupación debe constar de 2 columnas y la última debe constar de 1 columna.

Por tanto la matriz resultante será:

$$A = \begin{bmatrix} a_{11} & \vdots & a_{12} & a_{13} & \vdots & a_{14} \\ a_{21} & \vdots & a_{22} & a_{23} & \vdots & a_{24} \\ a_{31} & \vdots & a_{32} & a_{33} & \vdots & a_{34} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{41} & \vdots & a_{42} & a_{43} & \vdots & a_{44} \\ a_{51} & \vdots & a_{52} & a_{53} & \vdots & a_{54} \end{bmatrix}$$

$\square$

En base a la matriz del ejemplo anterior se tiene que:

$$A_{11} := A[1,2,3|1] = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}$$

$$A_{12} := A[1,2,3|2,3] = \begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$$

$$A_{13} := A[1,2,3|4] = \begin{pmatrix} a_{14} \\ a_{24} \\ a_{34} \end{pmatrix}$$

$$A_{21} := A[4,5|1] = \begin{pmatrix} a_{41} \\ a_{51} \end{pmatrix}$$

$$A_{22} := A[4,5|2,3] = \begin{bmatrix} a_{42} & a_{43} \\ a_{52} & a_{53} \end{bmatrix}$$

$$A_{23} := A[4,5|4] = \begin{pmatrix} a_{44} \\ a_{54} \end{pmatrix}$$

De este modo, la matriz particionada quedaría escrita de la siguiente forma.

$$A = \begin{bmatrix} A_{11} & \vdots & A_{12} & \vdots & A_{13} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{21} & \vdots & A_{22} & \vdots & A_{23} \end{bmatrix} =: \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix} \tag{1.11}$$

Notar ahora el siguiente ejemplo.

Ejemplo 1.4. *Dada la matriz $A = [a_{ij}]_{m\times n}$, determine sus submatrices y la partición de ella según el siguiente patrón de partición.*

$$\begin{cases} m = m_1 + m_2 \\ n = n_1 + n_2 \end{cases}$$

Solución:

De acuerdo a lo analizado en el ejemplo anterior se tiene que:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n_1} & \vdots & a_{1(n_1+1)} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n_1} & \vdots & a_{2(n_1+1)} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{m_1 1} & a_{m_1 2} & \cdots & a_{m_1 n_1} & \vdots & a_{m_1(n_1+1)} & \cdots & a_{m_1 n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{(m_1+1)1} & a_{(m_1+1)2} & \cdots & a_{(m_1+1)n_1} & \vdots & a_{(m_1+1)(n_1+1)} & \cdots & a_{m_1 n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn_1} & \vdots & a_{m(n_1+1)} & \cdots & a_{mn} \end{bmatrix}$$

De donde se obtienen las siguientes submatrices.

$$\begin{aligned} A_{11} &= A[1, 2, \ldots, m_1 | 1, 2, \ldots, n_1] \\ A_{12} &= A[1, 2, \ldots, m_1 | n_1 + 1, n_1 + 2, \ldots, n] \\ A_{21} &= A[m_1 + 1, m_1 + 2, \ldots, m | 1, 2, \ldots, n_1] \\ A_{21} &= A[m_1 + 1, m_1 + 2, \ldots, m | n_1 + 1, n_1 + 2, \ldots, n] \end{aligned}$$

Por lo tanto:

$$A = \begin{bmatrix} A_{11} & \vdots & A_{12} \\ \cdots & \cdots & \cdots \\ A_{21} & \vdots & A_{22} \end{bmatrix} =: \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

□

Dado el análisis realizado en los ejemplos anteriores, puede notarse que, si se considera una matriz $A_{m\times n}(\mathbb{R})$, al tomar el patrón de partición:

$$\begin{cases} m = m_1 + m_2 + \cdots + m_p \\ n = n_1 + n_2 + \cdots + n_q \end{cases}$$

Se obtendrá que $A = [A_{ij}]_{p\times q}$, lo cual en forma natural por lo antes visto, se escribirá de la siguiente manera.

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix}$$

Observación 1.18. *Se puede apreciar que en la matriz particionada anterior, y en la de los ejemplos anteriores, no se ha especificado cual es el orden de cada submatriz A_{ij}. En forma conveniente, cuando esto se quiera especificar, se usará la notación $A_{ij(r\times s)}$. Por decir en (1.11) se tendría lo siguiente.*

$$A = \begin{bmatrix} A_{11(3\times1)} & A_{12(2\times3)} & A_{13(3\times1)} \\ A_{21(2\times1)} & A_{22(2\times2)} & A_{23(2\times1)} \end{bmatrix}$$

Nota 1.21. *Las expresiones "submatriz" y "bloque" son sinónimas.*

1.10.3. Operaciones con matrices particionadas

Matrices equidimensionales

Definición 1.37. *Sea $A = [A_{ij}]_{p\times q}$ una matriz particionada y sean A_{ij}, A_{rs} un par de submatrices de A. Las matrices A_{ij}, A_{rs} son equidimensionales, si y sólo, si tienen la misma dimensión.*

Suma de matrices particionadas

Definición 1.38. *Sean $A = [A_{ij}]_{p\times q}$, $B = [B_{ij}]_{p\times q}$ un par de matrices particionadas, con la misma cantidad de bloques tales que, $dim(A_{ij}) = dim(B_{ij})$ para cualquier i, j. La suma de A con B, $A + B$, está dada por la matriz $[A_{ij} + B_{ij}]_{p\times q}$.*

Observación 1.19. *Claramente, la definición anterior, está diciendo que, si:*

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix} \quad , \quad B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1q} \\ B_{21} & B_{22} & \cdots & B_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ B_{p1} & B_{p2} & \cdots & B_{pq} \end{bmatrix}$$

y además:

$$\begin{aligned} \dim(A_{11}) &= \dim(B_{11}) \\ \dim(A_{12}) &= \dim(B_{12}) \\ &\vdots \\ \dim(A_{pq}) &= \dim(B_{pq}) \end{aligned}$$

entonces:

$$A+B=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1q} \\ B_{21} & B_{22} & \cdots & B_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ B_{p1} & B_{p2} & \cdots & B_{pq} \end{bmatrix}$$

$$:= \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \cdots & A_{1q}+B_{1q} \\ A_{21}+B_{21} & A_{22}+B_{22} & \cdots & A_{2q}+B_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{p1}+B_{p1} & A_{p2}+B_{p2} & \cdots & A_{pq}++B_{pq} \end{bmatrix}$$

Producto de una matriz particionada por un escalar

Definición 1.39. *Sea $A=[A_{ij}]_{p\times q}$ una matriz particionada y sea $\lambda \in (\mathbb{R})$ un escalar cualquiera. El producto de λ por la A, $\lambda\, A$, está dado por la matriz $[\lambda\, A_{ij}]_{p\times q}$.*

Observación 1.20. *La definición anterior, está diciendo que dado el número real λ y la submatriz $A=[A_{ij}]_{p\times q}$, se tiene que:*

$$\lambda\, A=\lambda \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix} := \begin{bmatrix} \lambda A_{11} & \lambda A_{12} & \cdots & \lambda A_{1q} \\ \lambda A_{21} & \lambda A_{22} & \cdots & \lambda A_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda A_{p1} & \lambda A_{p2} & \cdots & \lambda A_{pq} \end{bmatrix}$$

Multiplicación de matrices particionadas

Los casos que se estudiaran a continuación, han sido extraídos de [14] y son consecuencia inmediata de la definición de multiplicación de matrices.

Caso 1: Matriz - columna

Considerando que $A_{m\times n}(\mathbb{R})$ y $B_{n\times r}(\mathbb{R})$, por lo visto en la sección de *vectores asociados a una matriz*, se tiene que:

$$
\begin{aligned}
A\,B &= A\left[\begin{array}{cccc} Col_1(B) & Col_2(B) & \cdots & Col_r(B) \end{array}\right] \\
&= A\left[\begin{array}{ccccccc} Col_1(A) & \vdots & Col_2(A) & \vdots & \cdots & \vdots & Col_r(A) \end{array}\right] \\
&= \left[\begin{array}{ccccccc} A\,Col_1(A) & \vdots & A\,Col_2(A) & \vdots & \cdots & \vdots & A\,Col_r(A) \end{array}\right] \\
&= \left[\begin{array}{cccc} A\,Col_1(B) & A\,Col_2(B) & \cdots & A\,Col_r(B) \end{array}\right]
\end{aligned}
$$

Caso 2: Fila - matriz

Al igual que en el caso anterior, si se considera la matriz $A_{m\times n}(\mathbb{R})$ y la matriz $B_{n\times r}(\mathbb{R})$, por lo visto en la sección de *vectores asociados a una matriz*, se tiene que:

$$
A\,B = \begin{bmatrix} Fil_1(A) \\ Fil_2(A) \\ \vdots \\ Fil_m(A) \end{bmatrix} B
$$

Por lo tanto:

$$
A\,B = \begin{bmatrix} Fil_1(A) \\ \cdots \\ Fil_2(A) \\ \cdots \\ \vdots \\ \cdots \\ Fil_m(A) \end{bmatrix} B = \begin{bmatrix} Fil_1(A)B \\ \cdots \\ Fil_2(A)B \\ \cdots \\ \vdots \\ \cdots \\ Fil_m(A)B \end{bmatrix} = \begin{bmatrix} Fil_1(A)B \\ Fil_2(A)B \\ \vdots \\ Fil_m(A)B \end{bmatrix}
$$

en algún momento

Caso 3: Columna - fila

Teorema 1.44.
Si $A_{m\times n}(\mathbb{R})$ *y* $B_{n\times r}(\mathbb{R})$, *se satisface que:*

$$A\,B = \begin{bmatrix} Col_1(A) & Col_2(A) & \cdots & Col_n(A) \end{bmatrix} \begin{bmatrix} Fil_1(B) \\ Fil_2(B) \\ \vdots \\ Fil_n(B) \end{bmatrix}$$

$$= \begin{bmatrix} Col_1(A) & \vdots & Col_2(A) & \vdots & \cdots & \vdots & Col_n(A) \end{bmatrix} \begin{bmatrix} Fil_1(B) \\ \cdots \\ Fil_2(B) \\ \cdots \\ \vdots \\ \cdots \\ Fil_n(B) \end{bmatrix}$$

$$= Col_1(A)Fil_1(B) + Col_2(A)Fil_2(A) + \cdots Col_n(A)Fil_n(A)$$

Demostración: Revisar [9]

Obteniendo así, el siguiente esquema.

$$\boxed{AB}=\begin{cases} A\left[Col_1(B) \;\vdots\; Col_2(B) \;\vdots\; \cdots \;\vdots\; Col_r(B)\right]=\left[ACol_1(B) \;\vdots\; ACol_2(B) \;\vdots\; \cdots \;\vdots\; ACol_r(B)\right] \\[2em] \begin{bmatrix} Fil_1(A) \\ \cdots \\ Fil_2(A) \\ \cdots \\ \vdots \\ \cdots \\ Fil_n(A) \end{bmatrix} B = \begin{bmatrix} Fil_1(A)B \\ \cdots \\ Fil_2(A)B \\ \cdots \\ \vdots \\ \cdots \\ Fil_n(A)B \end{bmatrix} \\[2em] \left[Col_1(A) \;\vdots\; Col_2(A) \;\vdots\; \cdots \;\vdots\; Col_n(A)\right] \begin{bmatrix} Fil_1(B) \\ \cdots \\ Fil_2(B) \\ \cdots \\ \vdots \\ \cdots \\ Fil_n(B) \end{bmatrix} = Col_1(A)Fil_1(B)+Col_2(A)Fil_2(A)+\cdots Col_n(A)Fil_n(A) \end{cases}$$

Teorema 1.45.
Si $A_{p\times q}(\mathbb{R})$ y $B_{q\times r}(\mathbb{R})$ son matrices particionadas, se verifica que:

$$A\,B = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1r} \\ B_{21} & B_{22} & \cdots & B_{23} \\ \vdots & \vdots & \vdots & \vdots \\ B_{q1} & B_{q2} & \cdots & B_{qr} \end{bmatrix}$$

$$= \begin{bmatrix} \sum\limits_{j=1}^{q} A_{1j}B_{j1} & \sum\limits_{j=1}^{q} A_{1j}B_{j2} & \cdots & \sum\limits_{j=1}^{q} A_{1j}B_{jr} \\ \sum\limits_{j=1}^{q} A_{2j}B_{j1} & \sum\limits_{j=1}^{q} A_{2j}B_{j2} & \cdots & \sum\limits_{j=1}^{q} A_{2j}B_{jr} \\ \vdots & \vdots & \vdots & \vdots \\ \sum\limits_{j=1}^{q} A_{pj}B_{j1} & \sum\limits_{j=1}^{q} A_{pj}B_{j2} & \cdots & \sum\limits_{j=1}^{q} A_{pj}B_{jr} \end{bmatrix}$$

además:

$$A^T = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{bmatrix}^T = \begin{bmatrix} A_{11}^T & A_{21}^T & \cdots & A_{p1}^T \\ A_{12}^T & A_{22}^T & \cdots & A_{p2}^T \\ \vdots & \vdots & \vdots & \vdots \\ A_{1q}^T & A_{2q}^T & \cdots & A_{pq}^T \end{bmatrix}$$

Demostración: Revisar [12]

Nota 1.22. *El lector interesado en profundizar un poco más en esta sección, puede revisar cada una de las referencias que han sido mencionadas.*

1.11. Transformaciones lineales básicas

Transformación lineal

Definición 1.40. *Una aplicación $T:\mathbb{R}^m \to \mathbb{R}^n$ es una transformación lineal, si y sólo si:*

1. $\forall\, \overline{x}, \overline{y} \in \mathbb{R}^m : T(\overline{x}+\overline{y}) = T(\overline{x}) + T(\overline{y})$.
2. $(\forall\, \lambda \in \mathbb{R})(\forall\, \overline{x} \in \mathbb{R}^m) : T(\lambda\, \overline{x}) = \lambda\, T(\overline{x})$.

Nota 1.23. *Las expresiones "transformación lineal", "aplicación lineal" y "operador lineal" son sinónimas.*

Teorema 1.46.
Una aplicación $T : \mathbb{R}^m \to \mathbb{R}^n$ es una transformación lineal, si y sólo si:

$$(\forall\, \overline{x}\,,\, \overline{y} \in \mathbb{R}^m)(\forall\, \lambda \in \mathbb{R}) : T(\lambda\, \overline{x} + \overline{y}) = \lambda\, T(\overline{x}) + T(\overline{y}).$$

Demostración: Revisar [12] o [3]

Teorema 1.47.
Dada una matriz $A_{m\times n}(\mathbb{R})$, se verifica que la aplicación:

$$\begin{aligned} T_A : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ \overline{x} &\mapsto T_A(\overline{x}) := A\,\overline{x}\,,\ \forall\,\overline{x} \end{aligned}$$

es una transformación lineal.

Demostración: Revisar [14]

Nota 1.24. *La aplicación T_A del resultado anterior, es llamada transformación matricial.*

Teorema 1.48.
Si $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es una transformación lineal, entonces existe la matriz

$$A = \begin{bmatrix} T(e_1) & \vdots & T(e_2) & \vdots & \cdots & \vdots & T(e_n) \end{bmatrix}_{m\times n}$$

de modo que $T = T_A$

Demostración: Revisar [14]

Observación 1.21.

▷ *El resultado anterior, está diciendo que toda transformación lineal es una transformación matricial.*

▷ *La matriz $A := \begin{bmatrix} T(e_1) & \vdots & T(e_2) & \vdots & \cdots & \vdots & T(e_n) \end{bmatrix}_{m\times n}$ del teorema anterior, es llamada matriz estándar de la transformación lineal T.*

▷ *Usualmente la matriz estándar asociada una transformación lineal T, se denota por $[T]$. En este sentido se tendría que $\forall\, \overline{x} \in \mathbb{R}^n : T(\overline{x}) = [T]\, \overline{x}$.*

Teorema 1.49.
Si $T : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ y $S : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ son transformaciones lineales, entonces la aplicación $S \circ T : \mathbb{R}^m \longrightarrow \mathbb{R}^p$ es una transformación lineal. Además se verifica que $[S \circ T] = [S]\,[T]$.

Demostración: Revisar [14]

Teorema 1.50.
Si $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ y es una transformación lineal invertible, entonces la aplicación inversa de T, $T^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, también es una transformación lineal. Además se verifica que $[T^{-1}] = [T]^{-1}$.

Demostración: Revisar [14]

1.11.1. Espacios asociados a una transformación lineal

Teorema 1.51.
Dada una transformación lineal $T : \mathbb{R}^m \to \mathbb{R}^n$, se verifica que:

1. *Si U es un subespacio de $\mathbb{R}^m$, entonces el conjunto:*

$$T(U) = \{T(\overline{x})/\overline{x} \in U\} = \{\overline{y} \in \mathbb{R}^n/\overline{y} = T(\overline{x}) \text{ para algún } \overline{x} \in U\}$$

es un subespacio en $\mathbb{R}^n$.

2. *Si V es un subespacio de $\mathbb{R}^n$, entonces el conjunto:*

$$T^{-1}(V) = \{\overline{x} \in \mathbb{R}^m/T(\overline{x}) \in V\}$$

es un subespacio en $\mathbb{R}^m$.

Demostración: Revisar [12]

Imagen de una transformación lineal

Definición 1.41. *Sea $T : \mathbb{R}^m \to \mathbb{R}^n$ una transformación lineal. La imagen de T, $Im(T)$, está dada por $\{T(\overline{x})/\overline{x} \in \mathbb{R}^m\}$.*

Núcleo de una transformación lineal

Definición 1.42. *Sea $T : \mathbb{R}^m \to \mathbb{R}^n$ una transformación lineal. El núcleo de T, $Ker(T)$, está dada por $\{\overline{x} \in \mathbb{R}^m / T(\overline{x}) = \overline{0}\}$.*

Corolario 1.3.
Dada una transformación lineal $T : \mathbb{R}^m \to \mathbb{R}^n$, se verifica que:

1. *$Im(T) = T(\mathbb{R}^m)$ es un subespacio de $\mathbb{R}^n$.*
2. *$Ker(T) = T^{-1}\{\overline{0}\}$ es un subespacio de $\mathbb{R}^m$.*

Demostración: Revisar [3]

Rango y nulidad de una transformación lineal

Definición 1.43. *Sea $T : \mathbb{R}^m \to \mathbb{R}^n$ una transformación lineal. El rango de T, $rg(T)$, está dado por la dimensión de la imagen de T, i.e. $rg(T) := dim[Im(T)]$, la nulidad de T, $nul(T)$, está dada por la dimensión del núcleo de T, es decir $nul(T) := dim[Ker(T)]$.*

Teorema 1.52.
Sea $T : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ una transformación lineal. Si $S = gen\{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_k\}$, entonces $T(S) = gen\{T(\overline{u}_1), T(\overline{u}_2), \ldots, T(\overline{u}_k)\}$.

Demostración: Revisar [4]

Teorema 1.53.
Si $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es una transformación lineal y $A_{m\times n}(\mathbb{R})$ es la matriz estándar asociada, entonces se tiene que:

1. *$Im(T) = Col(A)$.*
2. *$rg(T) = rg(A)$.*
3. *$Ker(T) = Nul(A)$.*
4. *$dim[Ker(T)] =: nul(T) = dim[Nul(A)]$.*

Demostración: Revisar [12]

1.11.2. Operador proyección ortogonal

Aplicación proyección

Definición 1.44. *Sea W, un subespacio de $\mathbb{R}^n$. La aplicación*

$$\begin{aligned} Proy_W : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ \overline{x} &\mapsto Proy_W(\overline{x}) := \overline{p} \end{aligned}$$

tal que $\overline{x} - \overline{p} \perp W$, es llamada aplicación proyección ortogonal sobre W.

Teorema 1.54.
Siendo W un subespacio de $\mathbb{R}^n$, la aplicación:

$$\begin{aligned} Proy_W : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ \overline{x} &\mapsto Proy_W(\overline{x}) := \overline{p} \quad t.q.\ \overline{x} - \overline{p} \perp W. \end{aligned}$$

tiene las siguientes propiedades.

1. $Proy_W$ *es una transformación lineal.*
2. $\forall \overline{x} \in \mathbb{R}^n : Proy_W(\overline{x}) = \overline{x} \Leftrightarrow \overline{x} \in W.$
3. $\forall \overline{x} \in \mathbb{R}^n : Proy_W(\overline{x}) = \overline{0} \Leftrightarrow \overline{x} \in W^{\perp}.$
4. $\forall \overline{x} \in \mathbb{R}^n : Proy_W(Proy_W(\overline{x})) = Proy_W(\overline{x}).$
5. $Ker(Proy_W) = W^{\perp}$, $Im(Proy_W) = W.$

Demostración: Revisar [2]

Matriz de proyección ortogonal sobre W

Definición 1.45. *Sea W un subespacio de $\mathbb{R}^n$. La matriz $P_{n\times n}(\mathbb{R})$ que representa a la transformación lineal $Proy_W : \mathbb{R}^n \to \mathbb{R}^n$ en la base canónica, es decir la única matriz para cual se cumple que:*

$$Proy_W(\overline{x}) = P\,\overline{x}\,,\ \forall \overline{x} \in \mathbb{R}^n$$

es conocida como la matriz de proyección ortogonal sobre W.

Teorema 1.55.

Sea W un subespacio de $\mathbb{R}^n$. Si $P_{n\times n}(\mathbb{R})$ es la matriz de proyección ortogonal sobre W, entonces se tiene que:

1. $\forall\, \overline{x} \in \mathbb{R}^n : P\,\overline{x} = \overline{x} \Leftrightarrow \overline{x} \in W$.
2. $\forall\, \overline{x} \in \mathbb{R}^n : P\,\overline{x} = \overline{0} \Leftrightarrow \overline{x} \in W^{\perp}$.
3. $P^2 = P$, *es decir P es una matriz es idempotente.*
4. $Col(P) = W$, $Nul(P) = W^{\perp}$.
5. $I_n - P$ *es la matriz de proyección ortogonal sobre* $W^{\perp}$.
6. $P(I_n - P) = (I_n - P)P = \Theta$.
7. $Col(I_n - P) = W^{\perp} = Nul(P)$.
8. $Nul(I_n - P) = W = Col(P)$.
9. $P^T = P$, *es decir es simétrica.*

Demostración: Revisar [2]

Teorema 1.56.

Si $P_{n\times n}(\mathbb{R})$ es una matriz idempotente ($P^2 = P$) y simétrica ($P^T = P$), entonces P es una matriz de proyección ortogonal sobre su espacio columna, es decir:

$$\forall\, \overline{x} \in \mathbb{R}^n : P\,\overline{x} = Proy_{Col(P)}(\overline{x}).$$

Demostración: Revisar [2]

Capítulo 2

Inversa Generalizada de Moore-Penrose

Asumiendo que se tiene el siguiente problema:

$$\begin{cases} \quad A\overline{x} = \overline{b} \\ \text{S.a:} \\ \quad A \in \mathcal{M}_{m\times n}(\mathbb{R}) \\ \quad \overline{b} \in \mathbb{R}^m \\ \quad \overline{x} \in \mathbb{R}^n \ \text{(Vector incógnita)} \end{cases}$$

como se ha podido ver en el capítulo anterior, el teorema 1.20 es el único resultado del álgebra lineal clásica, que dice como sería la solución del problema anterior si se verificaran las condiciones establecidas en dicho resultado. Sin embargo, en general se sabe que la matriz A no tiene por qué ser cuadrada, y de serlo no tiene por qué ser invertible, bajo estas condiciones lo único que podría aplicarse es el teorema 1.21 y determinar si el problema anterior tiene o no solución.

Entonces bien, usando la teoría matricial estudiada en el capítulo anterior, dada la matriz $A_{m\times n}(\mathbb{R})$, este capítulo tiene los siguientes objetivos.

1. Construir en forma detallada la inversa generalizada de Moore-Penrose de dicha matriz A.
2. Garantizar la existencia y unicidad de dicha inversa.
3. Estudiar en detalle algunas propiedades de esta inversa
4. Dar una respuesta parcial y concreta del problema abierto mencionado en la introducción.

Esto con la finalidad de complementar el capítulo de las aplicaciones y dar respuesta al problema anterior.

2.1. Generalidades

Matriz simétrica definida positiva

Definición 2.1. *Una matriz simétrica $A_{n\times n}(\mathbb{R})$, es definida positiva, si y sólo si:*

$$\forall\, \overline{x} \in \mathbb{R}^n - \{\overline{0}\} : \overline{x}^{\,T} A\, \overline{x} > 0.$$

Matriz simétrica semidefinida positiva

Definición 2.2. *Una matriz simétrica $A_{n\times n}(\mathbb{R})$, es definida positiva, si y sólo si:*

$$\forall\, \overline{x} \in \mathbb{R}^n - \{\overline{0}\} : \overline{x}^{\,T} A\, \overline{x} \geq 0.$$

Teorema 2.1.
Si $A_{n\times n}(\mathbb{R})$ es una matriz simétrica, entonces para cualquier $\overline{x}$, $\overline{y} \in \mathbb{R}^n$, se tiene que:

1. $\langle \overline{x}, A\,\overline{y}\rangle = \langle A\,\overline{x}, \overline{y}\rangle$.

2. Para cualquier $m \in \mathbb{N}$, se verifica que A^m es una matriz simétrica.

3. $\langle \overline{x}, A^m\,\overline{y}\rangle = \langle A^m\,\overline{x}, \overline{y}\rangle$.

4. $\forall\, P(x) \in \mathbb{R}_m[x] : \langle \overline{x}, P(A)\,\overline{y}\rangle = \langle P(A)\,\overline{x}, \overline{y}\rangle$.

Demostración:

(*1*)
Teniendo en cuenta la observación 1.9, como:

$$\begin{aligned} A \in \mathcal{M}_{n\times n}(\mathbb{R}) &\Rightarrow \forall\, \overline{y} \in \mathbb{R}^n : (A\,\overline{y}) \in \mathbb{R}^n \\ &\Rightarrow \forall\, \overline{x} \in \mathbb{R}^n : \langle \overline{x}, A\overline{y}\rangle = \overline{x}^{\,T} \cdot A\,\overline{y} \\ &\qquad\qquad\qquad\qquad = (\overline{x}^{\,T} A) \cdot \overline{y} \\ &\qquad\qquad\qquad\qquad \overset{th.}{=} (A^T\,\overline{x})^T \cdot \overline{y} \\ &\qquad\qquad\qquad\qquad := (A\,\overline{x})^T \cdot \overline{y} \\ &\qquad\qquad\qquad\qquad = \langle A\,\overline{x}, \overline{y}\rangle \end{aligned}$$

Por lo tanto, es claro que:

$$\underline{\forall \overline{x}, \overline{y} \in \mathbb{R}^n : \langle A\,\overline{x}, \overline{y}\rangle = \langle \overline{x}, A\,\overline{y}\rangle}\Big|$$

(2)
Para probar este ítem debe hacerse uso del método de inducción.

1º) Para $k = 2$, se verifica que:

$$A^2 = A \cdot A := A^T \cdot A^T \overset{th.}{=} (A \cdot A)^T = (A^2)^T$$

2º) Asumir que para $k = m$, se verifica que $A^m = (A^m)^T$.

3º) Probar que para $k = m + 1$, se verifica que $A^k = (A^k)^T$.

Para este caso, notar que:

$$A^k = A^{m+1} = A^m \cdot A = (A^m)^T \cdot A := (A^m)^T \cdot A^T \overset{th.}{=} (A \cdot A^m)^T = (A^k)^T$$

Por lo tanto:

$$\forall\, m \in \mathbb{N} : A^m = (A^m)^T$$

(3)
Por *(1)* se sabe que, dada una matriz simétrica B, se verifica que:

$$\forall\, \overline{x}, \overline{y} \in \mathbb{R}^n : \langle \overline{x}, B\, \overline{y} \rangle = \langle B\, \overline{x}, \overline{y} \rangle$$

Luego, como por el ítem anterior se sabe que para todo m natural, A^m, es una matriz simétrica, es claro que:

$$\forall\, \overline{x}, \overline{y} \in \mathbb{R}^n : \langle \overline{x}, A^m\, \overline{y} \rangle = \langle A^m\, \overline{x}, \overline{y} \rangle$$

(4)
Como:

$$\begin{aligned} P(x) \in \mathbb{R}_m[x] &\Rightarrow P(x) = b_0 + b_1\, x + b_2\, x^2 + \cdots + b_m\, x^m \\ &\Rightarrow P(A) = b_0 + b_1\, A + b_2\, A^2 + \cdots + b_m\, A^m \\ &\Leftrightarrow P(A) = b_0\, I + b_1\, A + b_2\, A^2 + \cdots + b_m\, A^m \end{aligned}$$

Luego, como el conjunto de las matrices simétricas es un espacio vectorial, se tiene que $P(A)$ es también es una matriz simétrica y por lo tanto:

$$\forall\, \overline{x}, \overline{y} \in \mathbb{R}^n : \langle \overline{x}, P(A)\, \overline{y} \rangle = \langle P(A)\, \overline{x}, \overline{y} \rangle$$

■

Teorema 2.2.

Si $A_{n\times n}(\mathbb{R})$ es una matriz simétrica, entonces dos autovectores asociados a autovalores distintos de A son ortogonales.

Demostración:

Sean α, $\beta \in \mathbb{R}$ un par de autovalores distintos de A $(\alpha \neq \beta)$ y sean $\overline{x}$, $\overline{y} \in \mathbb{R}^n$ un par de vectores tales que:

$$\begin{aligned} \overline{x} \in E_\alpha(A) \quad &\wedge \quad \overline{y} \in E_\beta(A) \\ :\Rightarrow A\,\overline{x} = \alpha\,\overline{x} \quad &\wedge \quad A\,\overline{y} = \beta\,\overline{y} \\ :\Rightarrow A\,\overline{x} = \alpha\,\overline{x} \quad &\wedge \quad A^T\,\overline{y} = \beta\,\overline{y} \\ \overset{th.1,30}{\Rightarrow} \ \overline{x} \perp \overline{y} & \end{aligned}$$

■

Nota 2.1. *También es posible demostrar el resultado anterior, independizándose del teorema 1.30.*

Teorema 2.3.

Si $A_{n\times n}(\mathbb{R})$ es una matriz simétrica, entonces sus autovalores son reales.

Demostración:

Sea $\overline{x} \in \mathbb{R}^n - \{\overline{0}\}$ un autovector de A, con autovalor asociado λ, es decir: $A\,\overline{x} = \lambda\,\overline{x}$.

Ahora bien:

$$\langle A\,\overline{x}, \overline{x}\rangle = \langle \lambda\,\overline{x}, \overline{x}\rangle \overset{th.}{=} \lambda\,\langle \overline{x}, \overline{x}\rangle = \lambda||\overline{x}||^2 \tag{2.1}$$

Luego, como se sabe que $A\,\overline{x}$, $\overline{x} \in \mathbb{R}^n$, por definición de producto euclidiano se tiene que $\langle A\,\overline{x}, \overline{x}\rangle \in \mathbb{R}$ y $||\overline{x}||^2$ también es un número real, entonces de (2.1) se deduce que $\lambda \in \mathbb{R}$.

■

Teorema 2.4.

Si $A_{n\times n}(\mathbb{R})$ es una matriz semidefinida positiva, entonces todos sus autovalores son reales no negativos.

Demostración:

En primer lugar, como A es una matriz semidefinida positiva, por definición se sabe que es simétrica y por tanto todos sus autovalores son números reales.

Por otro lado, como A es semidefinida positiva, por definición se tiene que:

$$\forall\, \overline{x} \in \mathbb{R}^n : \overline{x}^{\,T} A\, \overline{x} \geq 0 \tag{2.2}$$

Ahora bien, si $\overline{y} \in \mathbb{R}^n - \{\overline{0}\}$ es un autovector de A, asociado a λ, por definición se tiene que:

$$\begin{aligned} A\,\overline{y} = \lambda\,\overline{y} \quad &\Rightarrow \quad \overline{y}^{\,T} A\,\overline{y} = \overline{y}^{\,T} \cdot (\lambda\,\overline{y}) = \langle \overline{y}, \lambda\,\overline{y} \rangle = \lambda\,\langle \overline{y}, \overline{y} \rangle = \lambda\,||\overline{y}||^2 \\ &\Rightarrow \quad \lambda\,||\overline{y}||^2 = \overline{y}^{\,T} A\,\overline{y} \underset{2.2}{\geq} 0 \\ &\Rightarrow \quad \lambda\,||\overline{y}||^2 \geq 0 \end{aligned} \tag{2.3}$$

Ahora bien, como $\overline{y} \neq \overline{0}$, entonces $||\overline{y}||^2 > 0$, por lo tanto de (2.3) se sigue que $\lambda \geq 0$.

■

Teorema 2.5.
Si $A_{n\times n}(\mathbb{R})$ es una matriz definida positiva, entonces todos sus autovalores son reales positivos.

Demostración:

Como A es definida positiva, entonces, por definición se sabe que A es simétrica y por tanto todos sus autovalores son números reales. Asimismo, como A es definida positiva, por definición se tiene que:

$$\forall\, \overline{x} \in \mathbb{R}^n - \{\overline{0}\} : \overline{x}^{\,T} A\, \overline{x} > 0 \tag{2.4}$$

Ahora bien, si $\overline{y} \in \mathbb{R}^n - \{\overline{0}\}$ es un autovector de A, asociado a λ, por definición se tiene que:

$$\begin{aligned} A\,\overline{y} = \lambda\,\overline{y} \quad &\Rightarrow \quad \overline{y}^{\,T} A\,\overline{y} = \overline{y}^{\,T} \cdot (\lambda\,\overline{y}) = \langle \overline{y}, \lambda\,\overline{y} \rangle = \lambda\,\langle \overline{y}, \overline{y} \rangle = \lambda\,||\overline{y}||^2 \\ &\Rightarrow \quad \lambda\,||\overline{y}||^2 = \overline{y}^{\,T} A\,\overline{y} \underset{2.4}{>} 0 \\ &\Rightarrow \quad \lambda\,||\overline{y}||^2 > 0 \end{aligned} \tag{2.5}$$

Ahora bien, como $\overline{y} \neq \overline{0}$, entonces $||\overline{y}||^2 > 0$, por lo tanto de (2.5) se sigue que $\lambda > 0$.

■

Teorema 2.6.
Si $A_{m\times n}(\mathbb{R})$, entonces $A^T\,A$ y $A\,A^T$ son matrices simétricas.

Demostración:

Si:
$$B := A^T\,A \quad \Rightarrow \quad B^T = (A^T\,A)^T = A^T\,(A^T)^T = A^T\,A := B$$

Concluyendo por tanto que $B := A^T\,A$ es una matriz simétrica. Análogamente, si:

$$B := A\,A^T \quad \Rightarrow \quad B^T = (A\,A^T)^T = (A^T)^T\,A^T = A\,A^T := B$$

Deduciendo por tanto que $B := A\,A^T$ es una matriz simétrica.

■

Observación 2.1.

▷ *Por definición se sabe que una matriz $A_{n\times n}(\mathbb{R})$ es semidefinida positiva si y sólo si:*
$$\forall\,\overline{x} \in \mathbb{R}^n : \overline{x}^{\,T} A\,\overline{x} \geq 0$$
$$\Leftrightarrow \forall\,\overline{x} \in \mathbb{R}^n : \overline{x}^{\,T} A\,\overline{x} = \langle \overline{x}, A\,\overline{x}\rangle \geq 0$$

▷ *Por definición se sabe que una matriz $A_{n\times n}(\mathbb{R})$ es semidefinida positiva si y sólo si:*
$$\forall\,\overline{x} \in \mathbb{R}^n - \{\overline{0}\} : \overline{x}^{\,T} A\,\overline{x} > 0$$
$$\Leftrightarrow \forall\,\overline{x} \in \mathbb{R}^n - \{\overline{0}\} : \overline{x}^{\,T} A\,\overline{x} = \langle \overline{x}, A\,\overline{x}\rangle > 0$$

Teorema 2.7.
Sea $A_{n\times n}(\mathbb{R})$ una matriz simétrica. La matriz A es semidefinida positiva si y sólo si todos sus autovalores son no negativos.

Demostración:

$(\Rightarrow)$
Esto es lo que se probo en el teorema 2.4.

$(\Leftarrow)$
Como A es una matriz simétrica, el teorema 1.43, garantiza que existe una base ortonormal de $\mathbb{R}^n$ formada por autovectores de A.

Si se asume que $\mathscr{B} := \{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_n\}$ es dicha base, entonces se tiene que:

$$\begin{cases} \forall i, j : \langle \overline{u}_i, \overline{u}_j \rangle = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases} \\ (\forall \overline{u}_j)(\exists \lambda_j \in \mathbb{R}) / A\overline{u}_j = \lambda_j \overline{u}_j \end{cases} \tag{2.6}$$

Ahora bien, debe probarse que A es una matriz semidefinida positiva, es decir debe probarse que:

$$\forall\, \overline{x} \in \mathbb{R}^n : \langle \overline{x}, A\, \overline{x} \rangle \geq 0$$

Por otro lado, como $\mathscr{B} := \{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_n\}$ es una base de $\mathbb{R}^n$, entonces:

$$\forall i \neq j : \overline{u}_i \neq \overline{u}_j \quad \overset{\text{Obs.1.13}}{\Rightarrow} \quad \lambda_i \neq \lambda_j$$

Ahora, dado $\overline{x} \in \mathbb{R}^n - \{\overline{0}\}$, como $\mathscr{B}$ es base, entonces:

$$\begin{aligned} \exists\, \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R} / \overline{x} = \alpha_1 \overline{u}_1 + \alpha_2 \overline{u}_2 + \cdots + \alpha_n \overline{u}_n \\ \Rightarrow \langle \overline{x}, A\overline{x} \rangle &= \left\langle \sum_{j=1}^{n} \alpha_j \overline{u}_j, A \sum_{k=1}^{n} \alpha_k \overline{u}_k \right\rangle \\ &= \left\langle \sum_{j=1}^{n} \alpha_j \overline{u}_j, \sum_{k=1}^{n} \alpha_k (A\overline{u}_k) \right\rangle \\ &\overset{(2.6)}{=} \left\langle \sum_{j=1}^{n} \alpha_j \overline{u}_j, \sum_{k=1}^{n} \alpha_k (\lambda_k \overline{u}_k) \right\rangle \\ &= \sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_j \alpha_k \lambda_k \langle \overline{u}_j, \overline{u}_k \rangle \end{aligned} \tag{2.7}$$

Luego, como para cualquier $i \neq j$ por (2.6) se sabe que $\langle \overline{u}_i, \overline{u}_j \rangle = 0$, al extender la doble sumatoria en (2.7), sólo permanecerán los sumandos en los cuales se verifique que $i = j$, pues el resto se anularán, es decir:

$$\langle \overline{x}, A\overline{x} \rangle = \sum_{j=1}^{n} \alpha_j \alpha_j \lambda_j \langle \overline{u}_j, \overline{u}_j \rangle \overset{(2.6)}{=} \sum_{j=1}^{n} \alpha_j^2 \lambda_j$$

Por último, como para cada j, λ_j es un autovalor de A, por hipótesis se tiene que $\lambda_j \geq 0$, asimismo, como para cada j, $\alpha_j^2 \geq 0$, se concluye que:

$$\begin{aligned} \forall j \in \overline{\{1, n\}} : \alpha_j^2 \lambda_j \geq 0 \quad \Rightarrow \quad \sum_{j=1}^{n} \alpha_j^2 \lambda_j \geq 0 \quad &\Rightarrow \quad \langle \overline{x}, A\overline{x} \rangle \geq 0\,,\ \forall \overline{x} \in \mathbb{R}^n - \{\overline{0}\} \\ &\Rightarrow \quad \langle \overline{x}, A\overline{x} \rangle \geq 0\,,\ \forall \overline{x} \in \mathbb{R}^n \end{aligned}$$

Concluyendo por tanto que A es una matriz semidefinida positiva.

■

Teorema 2.8.

Sea $A_{n\times n}(\mathbb{R})$ una matriz simétrica. La matriz A es definida positiva si y sólo si todos sus autovalores son positivos.

Demostración:

$(\Rightarrow)$

Esto es lo que se probo en el teorema 2.5.

$(\Leftarrow)$

Como A es una matriz simétrica, el teorema 1.43, garantiza que existe una base ortonormal de $\mathbb{R}^n$ formada por autovectores de A.

Si se asume que $\mathscr{B} := \{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_n\}$ es dicha base, entonces se tiene que:

$$\begin{cases} \forall i,j : \langle \overline{u}_i, \overline{u}_j \rangle = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases} \\ (\forall \overline{u}_j)(\exists \lambda_j \in \mathbb{R}) / A\overline{u}_j = \lambda_j \overline{u}_j \end{cases} \tag{2.8}$$

Ahora bien, debe probarse que A es una matriz definida positiva, es decir debe probarse que:

$$\forall \overline{x} \in \mathbb{R}^n - \{\overline{0}\} : \langle \overline{x}, A\,\overline{x} \rangle > 0$$

Por otro lado, como $\mathscr{B} := \{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_n\}$ es una base de $\mathbb{R}^n$, entonces:

$$\forall i \neq j : \overline{u}_i \neq \overline{u}_j \quad \overset{\text{Obs.1.13}}{\Rightarrow} \quad \lambda_i \neq \lambda_j$$

Ahora, dado $\overline{x} \in \mathbb{R}^n - \{\overline{0}\}$, como $\mathscr{B}$ es base, entonces:

$$\begin{aligned} &\exists\, \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R} / \overline{x} = \alpha_1 \overline{u}_1 + \alpha_2 \overline{u}_2 + \cdots + \alpha_n \overline{u}_n \\ \Rightarrow \langle \overline{x}, A\overline{x} \rangle &= \left\langle \sum_{j=1}^{n} \alpha_j \overline{u}_j, A \sum_{k=1}^{n} \alpha_k \overline{u}_k \right\rangle \\ &= \left\langle \sum_{j=1}^{n} \alpha_j \overline{u}_j, \sum_{k=1}^{n} \alpha_k (A\overline{u}_k) \right\rangle \\ &\overset{(2.8)}{=} \left\langle \sum_{j=1}^{n} \alpha_j \overline{u}_j, \sum_{k=1}^{n} \alpha_k (\lambda_k \overline{u}_k) \right\rangle \end{aligned}$$

Así, al aplicar las propiedades del producto interno se tiene que:

$$\langle \overline{x}, A\overline{x} \rangle = \sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_j \alpha_k \lambda_k \langle \overline{u}_j, \overline{u}_k \rangle \tag{2.9}$$

Luego, como para cualquier $i \neq j$ por (2.8) se sabe que $\langle \overline{u}_i, \overline{u}_j \rangle = 0$, al extender la doble sumatoria en (2.9), sólo permanecerán los sumandos en los cuales se verifique que $i = j$, pues el resto se anularán, es decir:

$$\langle \overline{x}, A\overline{x} \rangle = \sum_{j=1}^{n} \alpha_j \alpha_j \lambda_j \langle \overline{u}_j, \overline{u}_j \rangle \overset{(2.8)}{=} \sum_{j=1}^{n} \alpha_j^2 \lambda_j$$

Por último, como para cada j, λ_j es un autovalor de A, por hipótesis se tiene que $\lambda_j \geq 0$, asimismo, como para cada j, $\alpha_j^2 \geq 0$, se concluye que:

$$\forall j \in \overline{\{1, n\}} : \alpha_j^2 \lambda_j \geq 0 \quad \Rightarrow \quad \sum_{j=1}^{n} \alpha_j^2 \lambda_j \geq 0 \quad \Rightarrow \quad \langle \overline{x}, A\overline{x} \rangle \geq 0 \,,\ \forall \overline{x} \in \mathbb{R}^n - \{\overline{0}\}$$

Ahora bien, por lo antes concluido, la única posibilidad que hay para que $\langle \overline{x}, A\overline{x} \rangle$ sea cero, es que A sea la matriz nula, pero esto no puede ser, pues por hipótesis se tiene que los autovalores de A son todos positivos, es decir estrictamente mayores que cero y por la observación 1.13, se sabe que la matriz nula sólo tiene como autovalor al número cero, por tanto:

$$\langle \overline{x}, A\overline{x} \rangle > 0 \,,\ \forall \overline{x} \in \mathbb{R}^n - \{\overline{0}\}$$

Concluyendo por tanto que A es una matriz definida positiva.

■

Teorema 2.9.
Sea $A_{n\times n}(\mathbb{R})$ una matriz simétrica. Si A es una matriz semidefinida positiva, entonces existe una matriz $A^{1/2} \in \mathcal{M}_{n\times n}(\mathbb{R})$ simétrica tal que $A = A^{1/2}A^{1/2}$. Además, si A es definida positiva entonces A es invertible y existe una matriz $A^{-1/2} \in \mathcal{M}_{n\times n}(\mathbb{R})$ de modo que $A^{-1} = A^{-1/2}A^{-1/2}$ y $A^{1/2}A^{-1/2} = I = A^{-1/2}A^{1/2}$.

Demostración:

Parte 1:
Como A es una matriz simétrica, el teorema 1.40 garantiza que existe una matriz ortogonal $Q_{n\times n}(\mathbb{R})$, de modo que $D = Q^T A\, Q$ es una matriz diagonal formada por los autovalores de A, algunos de los cuales pueden repetirse.

Afirmación 1: *D es una matriz semidefinida positiva.*

En efecto, por hipótesis se sabe que A es una semidefinida positiva, esto implica que:

$$\forall\, \overline{y} \in \mathbb{R}^n : \langle \overline{y}, A\, \overline{y} \rangle \geq 0 \tag{2.10}$$

Ahora bien:

$$\begin{aligned} &\forall\, \overline{x} \in \mathbb{R}^n : \langle \overline{x}, D\, \overline{x} \rangle = \big\langle \overline{x}, Q^T A\, Q \overline{x} \big\rangle = \overline{x}^T \cdot (Q^T A\, Q\, \overline{x}) \overset{\text{th.}}{=} (\overline{x}^T Q^T) \cdot (A\, Q\, \overline{x}) \\ \overset{\text{th.}}{\Rightarrow} &\forall\, \overline{x} \in \mathbb{R}^n : \langle \overline{x}, D\, \overline{x} \rangle = (Q\, \overline{x})^T \cdot (A\, Q\, \overline{x}) = \left\langle \underline{Q\overline{x}}\, , A\, \underline{Q\overline{x}} \right\rangle \overset{(2.10)}{\geq} 0 \end{aligned}$$

Por lo tanto D es una matriz semidefinida positiva.

Afirmación 2: *Los elementos diagonales de D son no negativos.*

En efecto, como D es una matriz semidefinida positiva, el teorema 2.7 garantiza que los autovalores de D son no negativos, luego como D es una matriz diagonal sus autovalores justamente son sus elementos diagonales, verificándose por tanto la afirmación.

Por otro lado, si:

$$\begin{aligned} D := diag(d_{11}, d_{22}, \ldots, d_{nn}) &\Rightarrow \forall j \in \overline{\{1, n\}} : d_{jj} \geq 0 \\ &\Rightarrow \forall j \in \overline{\{1, n\}} : \exists \sqrt{d_{jj}} \geq 0 \\ &\Rightarrow \exists R := diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}}) \end{aligned} \tag{2.11}$$

Luego, como existen las matrices $Q, R \in \mathcal{M}_{n\times n}(\mathbb{R})$, se tiene que:

$$\exists\, A^{1/2} := Q\, R\, Q^T$$

Afirmación 3: *$A^{1/2}$ es una matriz simétrica.*

En efecto:

$$\begin{aligned} (A^{1/2})^T := \Big(\underline{QR}\, Q^T \Big)^T &\overset{\text{th.}}{=} (Q^T)^T (QR)^T \overset{\text{th.}}{=} QR^T Q^T \\ :\Rightarrow (A^{1/2})^T &= Q \Big[diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}}) \Big]^T Q^T \\ &\overset{\text{th.}}{=} Q\, diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}})\, Q^T \end{aligned}$$

De lo anterior se sigue en forma inmediata que $(A^{1/2})^T = QRQ^T := A^{1/2}$, concluyendo así que $A^{1/2}$ es una matriz simétrica.

Afirmación 4: *Se verifica que* $A^{1/2}A^{1/2} = A$.

En efecto:

$$\begin{aligned} A^{1/2}A^{1/2} &= (QRQ^T)(QRQ^T) \overset{\text{th.}}{=} (QR)(Q^TQ)(RQ^T) := QR(I)RQ^T \\ \Rightarrow A^{1/2}A^{1/2} &= QR^2Q^T \end{aligned} \tag{2.12}$$

Por otra parte, como por (2.11) se sabe que:

$$\begin{aligned} R^2 &= RR \\ &= diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}})\, diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}}) \\ &= diag(d_{11}, d_{22}, \ldots, d_{nn}) \\ &:= D \end{aligned} \tag{2.13}$$

Así, de (2.12) y (2.13) se sigue que:

$$A^{1/2}A^{1/2} = QDQ^T = Q(Q^TAQ)Q^T \overset{\text{th}}{=} (QQ^T)A(QQ^T) := A \tag{2.14}$$

Parte 2:

Afirmación 5: *A es una matriz invertible.*

En efecto, si se asume que A, es una matriz definida positiva, por ser A simétrica, realizando el mismo proceso de construcción que se realizó en la ***parte 1***, se tendrá que los elementos diagonales de D son estrictamente positivos, esto porque D tiene como elementos diagonales a los autovalores de A, y al ser A como se ha señalado, el teorema 2.8 garantiza que dichos autovalores son estrictamente positivos, es decir:

$$\begin{aligned} & D = diag(d_{11}, d_{22}, \ldots, d_{nn}) \quad \wedge \quad \forall j \in \overline{\{1, n\}} : d_{jj} > 0 \\ & \overset{\text{th.}}{\Rightarrow} \det(D) = d_{11} \cdot d_{22} \cdot \cdots \cdot d_{nn} > 0 \\ & \overset{\text{th.}}{\Rightarrow} \exists\, D^{-1} \end{aligned}$$

Luego como Q es una matriz ortogonal, entonces por definición Q es invertible, además D es invertible por lo que se acaba de probar, así A es el producto de matrices invertibles, pues por (2.14) se sabe que $A = QDQ^T$, se concluye por lo tanto que A es invertible.

Afirmación 6: *R es una matriz invertible.*

En efecto, como:

$$
\begin{aligned}
\forall j \in \overline{\{1,n\}} : d_{jj} > 0 \Rightarrow \forall j \in \overline{\{1,n\}} : \sqrt{d_{jj}} > 0 & \\
\Rightarrow det(R) &:= \det\left[diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}})\right] \\
&\overset{\text{th.}}{=} \sqrt{d_{11}} \cdot \sqrt{d_{22}} \cdots \sqrt{d_{nn}} > 0 \\
\overset{\text{th.}}{\Rightarrow} \exists\, R^{-1} & \\
\Rightarrow \exists\, A^{-1/2} := QR^{-1}Q^T &
\end{aligned} \tag{2.15}
$$

Afirmación 7: *Se verifica que:* $A^{-1/2}A^{-1/2} = A^{-1}$

En efecto:

$$
\begin{aligned}
A^{-1/2}A^{-1/2} &:= (QR^{-1}Q^T)(QR^{-1}Q^T) \overset{\text{th.}}{=} (QR^{-1})\,\underline{(Q^TQ)}(R^{-1}Q^T) \\
:\Rightarrow A^{-1/2}A^{-1/2} &= (QR^{-1})(R^{-1}Q^T) = Q(R^2)^{-1}Q^T = QD^{-1}Q^T \\
\Rightarrow A^{-1/2}A^{-1/2} &= Q(Q^TAQ)^{-1}Q^T := Q(Q^{-1}\,\underline{AQ})^{-1}Q^T \\
\Rightarrow A^{-1/2}A^{-1/2} &\overset{\text{th.}}{=} Q(AQ)^{-1}(Q^{-1})^{-1}Q^T \overset{\text{th.}}{=} \underline{Q(Q^{-1}}\,A^{-1})\,\underline{QQ^T} \\
:\Rightarrow A^{-1/2}A^{-1/2} &= A^{-1}
\end{aligned}
$$

Afirmación 8: *Se verifica que:* $A^{1/2}A^{-1/2} = I = A^{-1/2}A^{1/2}$

En efecto:

$$
A^{1/2}A^{-1/2} := (QR\,\underline{\ Q^T)(Q\ }\,R^{-1}Q^T) := (Q\,\underline{R)(R^{-1}}\,Q^T) := QQ^T := I_{n\times n}
$$

Asimismo:

$$
A^{-1/2}A^{1/2} := (QR^{-1}\,\underline{Q^T)(Q}\,RQ^T) := (Q\,\underline{\ R^{-1})(R\ }\,Q^T) := QQ^T := I_{n\times n}
$$

$\blacksquare$

Teorema 2.10.
Sea $A_{n\times n}(\mathbb{R})$ una matriz simétrica. Si A es una matriz semidefinida positiva, entonces existe una matriz $B_{n\times n}(\mathbb{R})$ tal que $A = BB^T$.

Demostración:

Como A es una matriz simétrica, el teorema 1.40 garantiza que existe una matriz ortogonal $Q_{n\times n}(\mathbb{R})$, tal que $D = Q^TAQ$ es una matriz diagonal formada por los autovalores de A, algunos de los cuales pueden repetirse.

Luego, si A es semidefinida positiva entonces el teorema 2.7 garantiza que los autovalores de A son no negativos. Por lo tanto, si:

$$
\begin{aligned}
D := diag(d_{11}, d_{22}, \ldots, d_{nn}) &\Rightarrow \forall j \in \overline{\{1,n\}} : d_{jj} \geq 0 \\
&\Rightarrow \forall j \in \overline{\{1,n\}} : \sqrt{d_{jj}} \geq 0 \\
&\Rightarrow \exists R := diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}}) \\
&\Rightarrow \exists B := QRQ
\end{aligned}
$$

Afirmación: *Se verifica que $BB^T = A$.*

En efecto:

$$
BB^T := (QRQ)(QRQ)^T \overset{\text{th.}}{=} (QR\ \underline{Q)(Q^T}\ R^TQ^T) := (Q\ \underline{R)(R^T}\ Q^T) \tag{2.16}
$$

Por otro lado:

$$
\begin{aligned}
RR^T &:= diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}})\Big[diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}})\Big]^T \\
&= diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}})\ diag(\sqrt{d_{11}}, \sqrt{d_{22}}, \ldots, \sqrt{d_{nn}}) \\
&= diag(d_{11}, d_{22}, \ldots, d_{nn}) \\
&= D
\end{aligned} \tag{2.17}
$$

Luego al usar (2.17) en (2.16) se tiene que:

$$
BB^T = QDQ^T = \underline{Q(Q^T}\ A\ \underline{Q)Q^T} := A
$$

$$
\Rightarrow \underline{BB^T = A|}
$$

■

Teorema 2.11.

Dada una matriz $A_{n\times n}(\mathbb{R})$ *se verifica que:*
Si A *es una matriz simétrica de rango* r*, entonces existe una matriz* $B_{n\times n}(\mathbb{R})$*, de modo que:*

$$B^T AB = \begin{bmatrix} I_{p\times p} & \Theta & \Theta \\ \Theta & -I_{q\times q} & \Theta \\ \Theta & \Theta & \Theta \end{bmatrix}$$

siendo $p+q=r$.

Demostración:

Como A es una matriz simétrica el teorema 1.40 garantiza que existe una matriz ortogonal $Q_{n\times n}(\mathbb{R})$ de modo que:

$$Q^T AQ = D := diag(d_{11}, d_{22}, \ldots, d_{nn}) / d_{jj} = \lambda_j$$

donde cada λ_j es un autovalor de A, además como A es una matriz simétrica el teorema 2.3 garantiza que todos sus autovalores son números reales, es decir:

$$\begin{aligned} &\forall j \in \overline{\{1,n\}} : d_{jj} = \lambda_j \in \mathbb{R} \\ &\overset{\text{th.}}{\Rightarrow} \forall j \in \overline{\{1,n\}} : d_{jj} = 0 \quad \veebar \quad d_{jj} \neq 0 \\ &\Rightarrow \exists R := diag(r_{jj}) / r_{jj} := \begin{cases} \dfrac{1}{\sqrt{d_{jj}}} & , \quad d_{jj} \neq 0 \\ \\ 1 & , \quad d_{jj} = 0 \end{cases} \end{aligned} \tag{2.18}$$

Queda por garantizada también la existencia de la matriz:

$$\begin{aligned} B := QR \quad \Rightarrow \quad B^T AB &:= (QR)^T A(QR) \\ &\overset{\text{th.}}{=} (R^T \underline{Q^T) A(Q} R) \\ &:= R^T DR \end{aligned} \tag{2.19}$$

Por otro lado, como por definición se sabe que:

$$\begin{aligned} R = diag(r_{11}, r_{22}, \ldots, r_{nn}) \Rightarrow R^T &= [diag(r_{11}, r_{22}, \ldots, r_{nn})]^T \\ &:= diag(r_{11}, r_{22}, \ldots, r_{nn}) \\ &:= R \end{aligned} \tag{2.20}$$

De este modo, al usar (2.20) en (2.19) se tiene que $B^T AB = RDR$, luego como D y R son matrices diagonales, el producto RDR también será una matriz diagonal, la que en sus elementos diagonales contiene obviamente contiene los números 1, 0 y -1, según lo definido en (2.18), así ordenando en forma adecuada los elementos diagonales de $B^T AB$ se obtiene el resultado buscado.

■

2.2. Descomposición en valores singulares (D.V.S)

Introducción

El teorema 1.40 muestra que toda matriz simétrica $A_{n\times n}(\mathbb{R})$ se puede factorizarse como $A = QDQ^T$ donde Q es una matriz ortogonal y D es una matriz diagonal que muestra los autovalores de A. Si por el contrario, la matriz A no es simétrica pero es diagonalizable entonces ella puede ser factorizada como $A = PDP^{-1}$ donde D es como antes pero P es simplemente una matriz invertible.

El problema con lo antes mencionado es que no toda matriz es diagonalizable, no toda matriz es simétrica y por último no toda matriz es cuadrada. En esta sección, por sorprendente que parezca se probará que toda matriz simétrica o no, cuadrada o no, tiene una factorización de la forma PDQ^T, donde misteriosamente P, Q son matrices ortogonales y D es una matriz diagonal. Este asombroso resultado es conocido como la *Descomposición en valores singulares (D.V.S)* y resulta ser una de las factorizaciones más importantes que existe.

Observación 2.2. *Una matriz $A_{n\times n}(\mathbb{R})$ a lo mucho puede tener n autovalores. En efecto, esto se debe a que el teorema 1.28 garantiza que los autovalores de A son todas las raíces de su polinomio característico $P_\lambda(A)$. Por otro lado el teorema 1.27 muestra que el polinomio característico asociado a la matriz $A_{n\times n}(\mathbb{R})$, es de grado n, luego el teorema fundamental del álgebra, garantiza que dicho polinomio a lo más tiene n raíces. Así la matriz $A_{n\times n}(\mathbb{R})$ tiene a lo sumo n raíces.*

Teorema 2.12.

Si $A_{m\times n}(\mathbb{R})$, *entonces se tiene que:*

1. $Ker(A) = Ker(A^T A)$.
2. $Ker(A^T) = Ker(AA^T)$
3. $rg(A) = rg(A^T A) = rg(AA^T)$
4. AA^T *y* $A^T A$ *son matrices simétricas.*
5. AA^T *y* $A^T A$ *son matrices semidefinidas positivas.*
6. $A^T A$ *es definida positiva si y sólo si* $rg(A) = n$.
7. AA^T *es definida positiva si y sólo si* $rg(A) = m$.

siendo $Ker(A) := Nul(A)$.

Demostración:

(*1*)

Por definición se sabe que:

$$Ker(A) := \{\overline{x} \in \mathbb{R}^n / A\,\overline{x} = \overline{0}\}$$
$$Ker(B := A^T A) := \{\overline{z} \in \mathbb{R}^n / B\,\overline{z} = \overline{0}\}$$

Ahora bien, por definición de igualdad de conjuntos se sabe que:

$$\begin{aligned} & Ker(A) = Ker(A^T A) \\ \Leftrightarrow\ & Ker(A) \subset Ker(A^T A) \quad \wedge \quad Ker(A) \supset Ker(A^T A) \end{aligned}$$

$(\subset)$

Notar que:

$$\begin{aligned} \forall\, \overline{x} \in \mathbb{R}^n : \overline{x} \in Ker(A) &:\Rightarrow A\overline{x} = \overline{0} \\ &\Rightarrow A^T(A\overline{x}) = A^T\overline{0} = \overline{0} \\ &\Rightarrow \underline{A^T A}\, \overline{x} = \overline{0} \\ &:\Rightarrow \overline{x} \in Ker(A^T A) \\ \therefore Ker(A) \subset Ker(A^T A) & \end{aligned}$$

$(\supset)$

Observar que:

$$\forall\, \overline{x} \in \mathbb{R}^n : \overline{x} \in Ker(A^T A) :\Rightarrow A^T A\overline{x} = \overline{0} \tag{2.21}$$

Ahora bien:

$$
\begin{aligned}
0 = \left\langle \overline{x}, \overline{0} \right\rangle \overset{(2.21)}{=} \left\langle \overline{x}, A^T A\overline{x} \right\rangle = \overline{x}^T \cdot (A^T A\overline{x}) = (\overline{x}^T A^T) \cdot (A\overline{x}) = (A\overline{x})^T \cdot (A\overline{x}) \\
= \langle A\overline{x}, A\overline{x} \rangle
\end{aligned}
$$

De lo anterior se deduce que:

$$\langle A\overline{x}, A\overline{x} \rangle = 0 :\Rightarrow A\overline{x} = \overline{0} :\Rightarrow \overline{x} \in Ker(A)$$

Por tanto se ha probado que:

$$
\begin{aligned}
&\forall\, \overline{x} \in \mathbb{R}^n : \overline{x} \in Ker(A^T A) \Rightarrow \overline{x} \in Ker(A) \\
&\Rightarrow Ker(A^T A) \subset Ker(A) \subset Ker(A^T A) \\
&\Rightarrow \underline{Ker(A^T A) = Ker(A)}\big|
\end{aligned}
$$

(2)

Se debe notar que en el caso anterior, se ha probado que dada una matriz $B_{p\times q}(\mathbb{R})$ se verifica que $Ker(B) = Ker(B^T B)$, entonces en forma particular, si se asume que $B_{p\times q} := A^T_{n\times m}$, por lo antes dicho se tendrá que:

$$Ker(A^T) := Ker(B) = Ker(B^T B) := Ker[(A^T)^T A^T] = Ker(AA^T)$$

(3)

Por el teorema del rango (th. 1.19), se verifica que dada la matriz $A_{m\times n}(\mathbb{R})$

$$
\begin{aligned}
\dim[Col(A)] + \dim[Ker(A)] = n :\Rightarrow rg(A) + \dim[Ker(A)] = n \\
\Rightarrow rg(A) = n - \dim[Ker(A)]
\end{aligned}
$$

Luego como se sabe que

$$
\begin{aligned}
Ker(A) = Ker(A^T A) &\Rightarrow \dim[Ker(A)] = \dim[Ker(A^T A)] \\
&\Rightarrow rg(A) = n - \dim[Ker(A^T A)] \\
&\Rightarrow rg(A) = rg(A^T A) \qquad (2.22)
\end{aligned}
$$

Asimismo, dada la matriz $A^T_{n\times m}(\mathbb{R})$, por lo antes justificado, se tiene lo siguiente.

$$
\begin{aligned}
\dim[Col(A^T)] + \dim[Ker(A^T)] = m :\Rightarrow rg(A^T) + \dim[Ker(A^T)] = m \\
\Rightarrow rg(A^T) = m - \dim[Ker(A^T)]
\end{aligned}
$$

Por otro lado, como se sabe que:

$$\begin{aligned} Ker(A^T) = Ker(AA^T) &\Rightarrow \dim[Ker(A^T)] = \dim[Ker(AA^T)] \\ &\Rightarrow rg(A^T) = m - \dim[Ker(AA^T)] \\ &\Rightarrow rg(A^T) = rg(AA^T) \end{aligned} \tag{2.23}$$

Ahora bien, como por el teorema 1.14 se sabe que $rg(A) = rg(A^T)$, entonces de (2.22) y (2.23) se deduce que:

$$rg(A) = rg(A^T) = rg(AA^T) = rg(A^T A)$$

(4)

Esto se demostró en el teorema 2.6

(5)

Parte 1:

Para probar que AA^T es una matriz semidefinida positiva, basta demostrar que sus autovalores son no negativos, tal y como lo asegura el teorema 2.7.

Si $\lambda \in \mathbb{R}$ es un autovalor cualquiera de AA^T, por definición se tiene que:

$$\exists \overline{x} \in \mathbb{R}^n - \{\overline{0}\} / AA^T\overline{x} = \lambda\overline{x}$$

Por otro lado:

$$\begin{aligned} &0 \leq ||A\overline{x}|| \leq ||A\overline{x}||^2 = \langle A\overline{x}, A\overline{x}\rangle = (A\overline{x})^T \cdot (A\overline{x}) = (\overline{x}^T A^T) \cdot (A\overline{x}) \\ \Rightarrow\ &0 \leq \overline{x}^T \cdot (A^T A\overline{x}) = \overline{x}^T \cdot (A^T A\overline{x}) = \left\langle \overline{x}, A^T A\overline{x}\right\rangle = \langle \overline{x}, \lambda\overline{x}\rangle = \lambda \langle \overline{x}, \overline{x}\rangle \\ \Rightarrow\ &0 \leq \lambda||\overline{x}||^2 \end{aligned}$$

Ahora bien, como se sabe que $\overline{x} \neq \overline{0}$, por definición se deduce que $||\overline{x}|| > 0$ y por tanto $||\overline{x}||^2 > 0$, luego, como $\lambda||\overline{x}||^2 \geq 0$ entonces es claro que $\lambda \geq 0$. Así, como λ era un autovalor arbitrario de AA^T, entonces por lo antes mencionado se tiene que AA^T es una matriz semidefinida positiva.

Parte 2:

Se puede observar que en la *Parte 1* se ha demostrado que dada una matriz $B_{p\times q}(\mathbb{R})$, se verifica que la matriz BB^T es semidefinida positiva, por tanto si de manera particular se considera que $B_{p\times q} := A^T_{n\times m}$, por lo antes mencionado se tiene que $A^T(A^T)^T = A^T A$ es una matriz semidefinida positiva.

(6)

$(\Rightarrow)$

Si A^TA es una matriz definida positiva, el teorema 2.8 garantiza que todos sus autovalores son estrictamente positivos, esto quiere decir que la matriz A^TA no tiene al número cero como autovalor, bajo esta condición, el teorema 1.25 garantiza que A^TA es una matriz invertible, lo que quiere decir que $rg(A^TA) = n$, pero como se ha demostrado que $rg(A^TA) = rg(A)$, se concluye que $rg(A) = n$.

$(\Leftarrow)$

Si se verifica que el $rg(A) = n$, entonces por lo antes mencionado $rg(A^TA) = n$, pero como $A^TA \in \mathcal{M}_{n\times n}(\mathbb{R})$, el teorema 1.20 garantiza que A^TA es una matriz invertible, pero si esto sucede, el teorema 1.25 garantiza que, el número real cero no es un autovalor de A^TA y como por el ítem *5* se sabe que A^TA es una matriz semidefinida positiva, entonces A^TA es una matriz definida positiva.

(7)

Siguiendo el razonamiento utilizado en el ítem *6*, se tiene que:

$$\begin{aligned}
AA^T\text{es definida positiva} &\Leftrightarrow \text{Todos sus autovalores son positivos}\\
&\Leftrightarrow AA^T\text{no tiene al cero como autovalor}\\
&\Leftrightarrow AA^T\text{es una matriz invertible}\\
&\Leftrightarrow rg(AA^T) = m\\
&\Leftrightarrow rg(A) = m
\end{aligned}$$

■

Teorema 2.13.
Si $A_{m\times n}(\mathbb{R})$, entonces se tiene que:

1. *A^TA y AA^T tienen los mismos autovalores no nulos.*
2. *Si $\overline{x} \in \mathbb{R}^n - \{\overline{0}\}$ es un autovector de A^TA asociado a $\lambda \in \mathbb{R} - \{0\}$ entonces $A\overline{x}$ es un autovector de AA^T asociado a λ.*
3. *Si $\overline{y} \in \mathbb{R}^m - \{\overline{0}\}$ es un autovector de AA^T asociado a $\lambda \in \mathbb{R} - \{0\}$ entonces $A^T\overline{y}$ es un autovector de A^TA asociado a λ.*
4. *La multiplicidad de los autovalores no nulos de A^TA coincide con la de los de AA^T*

Demostración:

(*1*)

Parte 1:

Se debe probar que:

$$\forall \lambda \in \mathbb{R} - \{0\} : \lambda \in Sp(A^T A) \Rightarrow \lambda \in Sp(AA^T)$$

Por definición se sabe que:

$$\forall\, \lambda \in \mathbb{R} - \{0\} : \lambda \in Sp(A^T A) \Rightarrow \exists \overline{x} \in \mathbb{R}^n - \{\overline{0}\} / A^T A\overline{x} = \lambda\overline{x}$$

Ahora bien, como:

$$\begin{aligned} A^T A\overline{x} = \lambda\overline{x} \quad &\Rightarrow \quad A(A^T A\overline{x}) = A(\lambda\overline{x}) \\ &\Rightarrow \quad (AA^T)A\overline{x} = \lambda(A\overline{x}) \\ &\Rightarrow \quad \exists \overline{y} := A\overline{x} \in \mathbb{R}^m / AA^T\overline{y} = \lambda\overline{y} \\ &:\Rightarrow \quad \lambda \in Sp(AA^T) \end{aligned}$$

Afirmación 1: $\overline{y} := A\overline{x} \neq \overline{0}$

En efecto, si:

$$\begin{aligned} \overline{y} = \overline{0} \quad \Rightarrow \quad A\overline{x} = \overline{0} \quad &\Rightarrow \quad A = \Theta \quad \vee \quad \underbrace{\overline{x} = \overline{0}}_{F} \\ &\Rightarrow \quad A = \Theta \\ &\Rightarrow \quad A^T A = A^T\Theta = \Theta \\ &\overset{\text{Obs.1.13}}{\Rightarrow} \quad \lambda = 0 \quad (\Rightarrow\Leftarrow) \end{aligned}$$

La contradicción se obtiene porque desde un inicio se asumió que λ era no nulo.

Parte 2:

Se debe probar que:

$$\forall \lambda \in \mathbb{R} - \{0\} : \lambda \in Sp(A^T A) \Rightarrow \lambda \in Sp(AA^T)$$

Por definición se sabe que:

$$\forall\, \lambda \in \mathbb{R} - \{0\} : \lambda \in Sp(AA^T) \Rightarrow \exists \overline{x} \in \mathbb{R}^m - \{\overline{0}\} / AA^T\overline{x} = \lambda\overline{x}$$

Luego, como:

$$\begin{aligned} AA^T\overline{x} = \lambda\overline{x} \quad &\Rightarrow \quad A^T(AA^T\overline{x}) = A^T(\lambda\overline{x}) \\ &\Rightarrow \quad (A^TA)A^T\overline{x} = \lambda(A^T\overline{x}) \\ &\Rightarrow \quad \exists\overline{z} := A^T\overline{x} \in \mathbb{R}^m / A^TA\overline{z} = \lambda\overline{z} \\ &:\Rightarrow \quad \lambda \in Sp(A^TA) \end{aligned}$$

Afirmación 2: $\overline{z} := A^T\overline{x} \neq \overline{0}$

En efecto, si:

$$\begin{aligned} \overline{z} = \overline{0} \quad \Rightarrow \quad A^T\overline{x} = \overline{0} \quad &\Rightarrow \quad A^T = \Theta \quad \vee \quad \underbrace{\overline{x} = \overline{0}}_{F} \\ &\Rightarrow \quad A^T = \Theta \\ &\Rightarrow \quad AA^T = A\Theta = \Theta \\ &\overset{\text{Obs.1.13}}{\Rightarrow} \quad \lambda = 0 \quad (\Rightarrow\Leftarrow) \end{aligned}$$

La contradicción se obtiene porque desde un inicio se asumió que λ era no nulo.

En conclusión A^TA y AA^T tienen los mismos autovalores no nulos.

(2)

Sean $\overline{x} \in \mathbb{R}^n - \{\overline{0}\}$ y $\lambda \in \mathbb{R} - \{0\}$ un par de elementos tales que:

$$A^TA\overline{x} = \lambda\overline{x} \tag{2.24}$$

Afirmación 3: *Si* $\overline{y} := A\overline{x}$ *entonces* $AA^T\overline{y} = \lambda\overline{y}$

En efecto, si $\overline{y} := A\overline{x}$ se verifica que:

$$\begin{aligned} &AA^T\overline{y} = AA^T(A\overline{x}) = A(A^TA\overline{x}) \overset{(2.24)}{=} A(\lambda\overline{x}) = \lambda(A\overline{x}) = \lambda\overline{y} \\ \Rightarrow\ &AA^T\overline{y} = \lambda\overline{y} \end{aligned}$$

Lo que obviamente quiere decir que $\overline{y} := A\overline{x}$ es un autovector de AA^T, asociado a λ.

(3)

Sean $\overline{y} \in \mathbb{R}^m - \{\overline{0}\}$ y $\lambda \in \mathbb{R} - \{0\}$ un par de elementos tales que:

$$AA^T\overline{y} = \lambda\overline{y} \tag{2.25}$$

Afirmación 4: *Si* $\overline{z} := A^T\overline{y}$ *entonces* $A^TA\overline{z} = \lambda\overline{z}$

En efecto, si $\overline{z} := A^T\overline{y}$ se verifica que:

$$\begin{aligned} & A^TA\overline{z} = A^TA(A^T\overline{y}) = A^T(AA^T\overline{y}) \overset{(2.22)}{=} A^T(\lambda\overline{y}) = \lambda(A^T\overline{y}) = \lambda\overline{z} \\ \Rightarrow\ & A^TA\overline{z} = \lambda\overline{z} \end{aligned}$$

Lo que obviamente quiere decir que $\overline{z} := A^T\overline{y}$ es un autovector de A^TA, asociado a λ.

(4)

Debe tenerse claro que se desea demostrar que dado $\lambda \in \mathbb{R} - \{0\}$ se verifica que $ma(\lambda_{AA^T}) = ma(\lambda_{A^TA})$.

Entonces bien, como AA^T y A^TA son matrices simétricas, el teorema 1.41 asegura que ambas matrices son diagonalizables, y bajo esta última condición el teorema 1.37 afirma que:

$$\begin{cases} ma(\lambda_{AA^T}) = \dim[E_\lambda(AA^T)] \\ ma(\lambda_{A^TA}) = \dim[E_\lambda(A^TA)] \end{cases}$$

Por consiguiente, basta demostrar que $\dim[E_\lambda(AA^T)] = \dim[E_\lambda(A^TA)]$, para esto notar que:

Afirmación 5: *Si* $\overline{x}$, $\overline{y}$ *son vectores L.I en* $E_\lambda(A^TA)$ *entonces* $A\overline{x}$, $A\overline{y}$ *son vectores L.I en* $E_\lambda(AA^T)$.

En efecto:

$$\forall \overline{x}, \overline{y} \in \mathbb{R}^n - \{\overline{0}\} : \begin{cases} \overline{x} \in E_\lambda(A^TA) & \overset{(2)}{\Rightarrow} & A\overline{x} \in E_\lambda(AA^T) \\ \overline{y} \in E_\lambda(A^TA) & \overset{(2)}{\Rightarrow} & A\overline{y} \in E_\lambda(AA^T) \end{cases} \tag{2.26}$$

Ahora bien, si $\overline{x}$, $\overline{y}$ son vectores L.I en $E_\lambda(A^TA)$, considerando (2.26), se tiene que:

$$\begin{cases} A^TA\overline{x} = \lambda\overline{x} \\ A^TA\overline{y} = \lambda\overline{y} \\ \forall \alpha\overline{x} + \beta\overline{y} : \alpha\overline{x} + \beta\overline{y} = \overline{0} \Rightarrow \alpha = \beta = 0 \end{cases} \tag{2.27}$$

Se debe demostrar que $A\overline{x}$, $A\overline{y}$ son vectores L.I en $E_\lambda(AA^T)$. Para esto, observar que, si α_0 y β_0 son números reales cualesquiera tales que:

$$\begin{aligned}
\overline{0} = \alpha_0 A\overline{x} + \beta_0 A\overline{y} \quad &\Rightarrow \quad A^T\overline{0} = A^T(\alpha_0 A\overline{x}) + A^T(\beta_0 A\overline{y}) \\
&\Rightarrow \quad \overline{0} = \alpha_0(A^TA\overline{x}) + \beta_0(A^TA\overline{y}) \\
&\qquad \overset{(2.27)}{=} \alpha_0(\lambda\overline{x}) + \beta_0(\lambda\overline{y}) \\
&\qquad = \lambda(\alpha_0\overline{x} + \beta_0\overline{y}) \\
&\Rightarrow \underbrace{\lambda = 0}_{F} \quad \vee \quad \alpha_0\overline{x} + \beta_0\overline{y} = \overline{0} \\
&\Rightarrow \alpha_0\overline{x} + \beta_0\overline{y} = \overline{0} \\
&\overset{(2.27)}{\Rightarrow} \alpha_0 = \beta_0 = 0
\end{aligned}$$

Esto quiere decir que $A\overline{x}$, $A\overline{y}$ efectivamente son vectores L.I en $E_\lambda(AA^T)$.

Afirmación 6: *Si $\overline{x}$, $\overline{y}$ generan $E_\lambda(A^TA)$, entonces $A\overline{x}$, $A\overline{y}$ generan $E_\lambda(AA^T)$.*

Por la afirmación anterior ya se sabe que $A\overline{x}$, $A\overline{y}$ son vectores L.I en $E_\lambda(AA^T)$, resta probar entonces que todos vector de $E_\lambda(AA^T)$ es combinación lineal de $A\overline{x}$, $A\overline{y}$.

Para ello hay que notar lo siguiente, si $\overline{x}$, $\overline{y}$ generan $E_\lambda(A^TA)$, entonces se tiene lo siguiente:

$$\begin{cases}
[\forall\overline{a} \in E_\lambda(A^TA)](\exists\alpha_1, \beta_1 \in \mathbb{R})/\overline{a} = \alpha_1\overline{x} + \beta_1\overline{y} \\
\overline{x} \in E_\lambda(A^TA) \Rightarrow A^TA\overline{x} = \lambda\overline{x} \\
\overline{y} \in E_\lambda(A^TA) \Rightarrow A^TA\overline{y} = \lambda\overline{y}
\end{cases} \tag{2.28}$$

Ahora bien:

$$\begin{aligned}
&\forall\overline{b} \in \mathbb{R}^m - \{\overline{0}\} : \overline{b} \in E_\lambda(AA^T) \overset{(3)}{\Rightarrow} A^T\overline{b} \in E_\lambda(A^TA) \overset{(2.28)}{\Rightarrow} A^T\overline{b} = \alpha_1\overline{x} + \beta_1\overline{y} \\
&\Rightarrow \lambda A^T\overline{b} = \lambda(\alpha_1\overline{x}) + \lambda(\beta_1\overline{y}) = \alpha_1(\lambda\overline{x}) + \beta_1(\lambda\overline{y}) \overset{(2.28)}{=} \alpha_1(A^TA\overline{x}) + \beta_1(A^TA\overline{y}) \\
&\Rightarrow \lambda A^T\overline{b} = A^T(\alpha_1 A\overline{x} + \beta_1 A\overline{y}) \\
&\Rightarrow A^T[\lambda\overline{b} - (\alpha_1 A\overline{x} + \beta_1 A\overline{y})] = \overline{0} \\
&\Rightarrow \underbrace{A^T = \Theta}_{F} \quad \vee \quad \lambda\overline{b} - (\alpha_1 A\overline{x} + \beta_1 A\overline{y}) = \overline{0} \\
&\Rightarrow \lambda\overline{b} = \alpha_1 A\overline{x} + \beta_1 A\overline{y} \\
&\Rightarrow \overline{b} = \left(\frac{\alpha_1}{\lambda}\right) A\overline{x} + \left(\frac{\beta_1}{\lambda}\right) A\overline{y}
\end{aligned}$$

Esto quiere decir que $\exists \alpha := \dfrac{\alpha_1}{\lambda}, \beta := \dfrac{\beta_1}{\lambda}$ de modo que $\overline{b} = \alpha A\overline{x} + \beta A\overline{y}$, es decir, efectivamente $A\overline{x}$, $A\overline{y}$ son una base de $E_\lambda(AA^T)$.

Puede notarse que básicamente se ha demostrado que, si:

$$\dim[E_\lambda(A^TA)] = 2 \quad \Rightarrow \dim[E_\lambda(AA^T)] = 2 \quad \Rightarrow \dim[E_\lambda(A^TA)] = \dim[E_\lambda(AA^T)]$$

De este modo, siguiendo el mismo proceso de demostración de la afirmación 5 y 6, se tiene que, si:

$$\dim[E_\lambda(A^TA)] = k \quad \Rightarrow \dim[E_\lambda(AA^T)] = k \quad \Rightarrow \underline{\dim[E_\lambda(A^TA)] = \dim[E_\lambda(AA^T)]}|$$

■

2.3. Valores singulares de una matriz

Introducción

El teorema que acaba de probarse es trascendental para el objetivo de este capítulo, pues como se ha podido observar, dada una matriz $A_{m\times n}(\mathbb{R})$, la matriz $AA^T_{n\times n}(\mathbb{R})$ es simétrica y como consecuencia de ello el teorema 1.41 garantiza que dicha matriz es diagonalizable, más aún, el ítem *5* del teorema 2.12 garantiza que $AA^T_{n\times n}$ es una matriz semidefinida positiva, por tanto sus autovalores son números reales no negativos, tal y como se vio en el teorema 2.7.

Por otro lado, como $AA^T_{n\times n}(\mathbb{R})$ es una matriz diagonalizable, el teorema 1.42 acredita que dicha matriz tiene n autovalores no necesariamente diferentes, pero por lo antes dicho si son no negativos.

Dada la situación antes mencionada, es posible reenumerar dichos autovalores de modo que puedan ser escritos en forma no creciente, es decir:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

Ahora bien, dado que $\sqrt{\cdot} : \mathbb{R}_0^+ \to \mathbb{R}_0^+$, $x \mapsto \sqrt{x}$ es una función creciente, a partir de la desigualdad múltiple antes señalada, se tiene que:

$$\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \cdots \geq \sqrt{\lambda_n} \geq 0$$

Por tanto tiene sentido trabajar con las raíces cuadradas de los autovalores ordenados de la matriz $AA^T_{n\times n}(\mathbb{R})$, obviamente el mismo análisis puede realizarse con la matriz $A^TA_{n\times n}(\mathbb{R})$ y el resultado sería exactamente el mismo porque ambas matrices tienen los mismos autovalores no nulos, según lo que se probó en el primer ítem del teorema 2.13.

Valor singular de una matriz

Definición 2.3. *Sea $A_{m\times n}(\mathbb{R})$, sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ los n autovalores de la matriz $A^TA_{n\times n}(\mathbb{R})$ colocados en forma no creciente y sea $j \in \overline{\{1,n\}}$. El número real σ_j es un valor singular de A, si y sólo si $\sigma_j = \sqrt{\lambda_j}$.*

Observación 2.3. *Asumiendo que $\lambda_1, \lambda_2, \ldots, \lambda_n$ son los autovalores de $A^TA_{n\times n}(\mathbb{R})$ colocados en forma no creciente.*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

La definición anterior está diciendo lo siguiente.

$$\begin{array}{ccccccccc}
\lambda_1 & \geq & \lambda_2 & \geq & \cdots & \geq & \lambda_n & \geq & 0 \\
\Updownarrow & & \Updownarrow & & & & \Updownarrow & & \\
\sqrt{\lambda_1} & \geq & \sqrt{\lambda_2} & \geq & \cdots & \geq & \sqrt{\lambda_n} & \geq & 0 \\
\shortparallel & & \shortparallel & & & & \shortparallel & & \\
\sigma_1 & \geq & \sigma_2 & \geq & \cdots & \geq & \sigma_n & \geq & 0
\end{array}$$

Teorema 2.14.

Sea $A_{m\times n}(\mathbb{R})$ y $\{\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_n\}$ una base ortonormal de $\mathbb{R}^n$ formada por autovectores de la matriz $A^TA_{n\times n}$, acomodada de forma que sus autovalores asociados sean tales que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

Si A tiene r valores singulares no nulos, entonces $\{A\overline{v}_1, A\overline{v}_2, \ldots, A\overline{v}_r\}$ es una base ortogonal del espacio $Col(A)$ y además $rg(A) = r$.

Demostración:

Como por hipótesis se tiene que $\{\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_n\}$ una base ortonormal de $\mathbb{R}^n$, entonces se verifica que:

$$\begin{cases} \forall i,j \in \overline{\{1,n\}} : \langle \overline{v}_i, \overline{v}_j\rangle = \begin{cases} 1 & , \quad i = j \\ 0 & , \quad i \neq j \end{cases} \\ \\ (\forall \overline{a} \in \mathbb{R}^n)(\exists \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}) / \overline{a} = \alpha_1\overline{v}_1 + \alpha_2\overline{v}_2 + \cdots + \alpha_n\overline{v}_n \end{cases} \tag{2.29}$$

Asimismo, por hipótesis se tiene que $\{\lambda_1, \lambda_2, \ldots, \lambda_n\} \subset Sp(A)$ satisface las siguientes condiciones:

$$\begin{cases} \lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0 \\ \forall j \in \overline{\{1, n\}} : A^T A\overline{v}_j = \lambda_j \overline{v}_j \end{cases} \tag{2.30}$$

Con todo esto por lo visto en la observación 2.3 se tiene que:

$$\begin{array}{ccccccccc} \lambda_1 & \geq & \lambda_2 & \geq & \cdots & \geq & \lambda_n & \geq & 0 \\ \Updownarrow & & \Updownarrow & & & & \Updownarrow & & \\ \sqrt{\lambda_1} & \geq & \sqrt{\lambda_2} & \geq & \cdots & \geq & \sqrt{\lambda_n} & \geq & 0 \\ \shortparallel & & \shortparallel & & & & \shortparallel & & \\ \sigma_1 & \geq & \sigma_2 & \geq & \cdots & \geq & \sigma_n & \geq & 0 \end{array} \tag{2.31}$$

Ahora bien, si la matriz $A_{m\times n}(\mathbb{R})$ tiene sus r primeros valores singulares no nulos, se tendría que:

$$\begin{array}{ccccccccc} \lambda_{r+1} & = & \lambda_{r+2} & = & \cdots & = & \lambda_n & = & 0 \\ \sigma_{r+1} & = & \sigma_{r+2} & = & \cdots & = & \sigma_n & = & 0 \end{array} \tag{2.32}$$

Por lo tanto (2.31) quedaría expresado de la siguiente manera.

$$\begin{array}{ccccccccc} \lambda_1 & \geq & \lambda_2 & \geq & \cdots & \geq & \lambda_r & > & 0 \\ \Updownarrow & & \Updownarrow & & & & \Updownarrow & & \\ \sqrt{\lambda_1} & \geq & \sqrt{\lambda_2} & \geq & \cdots & \geq & \sqrt{\lambda_r} & > & 0 \\ \shortparallel & & \shortparallel & & & & \shortparallel & & \\ \sigma_1 & \geq & \sigma_2 & \geq & \cdots & \geq & \sigma_r & > & 0 \end{array} \tag{2.33}$$

Afirmación 1: $\{A\overline{v}_1, A\overline{v}_2, \ldots, A\overline{v}_r\} \subset Col(A)$

En efecto, para cualquier $j \in \overline{\{1, n\}}$ se tiene que:

$$A\overline{v}_j = \left[\begin{array}{ccccccc} Col_1(A) & \vdots & Col_2(A) & \vdots & \cdots & \vdots & Col_n(A) \end{array}\right] \left[\begin{array}{c} v_j(1) \\ \cdots \\ v_j(2) \\ \cdots \\ \vdots \\ \cdots \\ v_j(n) \end{array}\right]$$

Donde al aplicar el teorema 1.44 se tiene que:

$$\begin{aligned}
A\overline{v}_j &= v_j(1)Col_1(A) + v_j(2)Col_2(A) + \cdots + v_j(n)Col_n(A)\\
&\in gen\{Col_1(A), Col_2(A), \ldots, Col_n(A)\}\\
:\Rightarrow\ & A\overline{v}_j \in Col(A), \forall j \in \overline{\{1,n\}}\\
\Rightarrow\ & \underline{\{A\overline{v}_1, A\overline{v}_2, \ldots, A\overline{v}_n\} \subset Col(A)|}
\end{aligned}$$

Afirmación 2: $\forall i \neq j : \overline{v}_i \perp \lambda_j \overline{v}_j$

En efecto, para cualquier i distinto de j por (2.29) se tiene que:

$$\begin{aligned}
\langle \overline{v}_i, \overline{v}_j\rangle = 0 \quad \Rightarrow \quad \lambda_j \langle \overline{v}_i, \overline{v}_j\rangle = \lambda_j 0 = 0 \quad \Rightarrow \quad \langle \overline{v}_i, \lambda_j\overline{v}_j\rangle = 0\\
:\Rightarrow \quad \underline{\overline{v}_i \perp \lambda_j\overline{v}_j|}
\end{aligned}$$

Afirmación 3: $\forall j \in \overline{\{1,n\}} : ||A\overline{v}_j|| = \sqrt{\lambda_j} := \sigma_j$

En efecto:

$$\begin{aligned}
&\forall j \in \overline{\{1,n\}} : ||A\overline{v}_j||^2 = \langle A\overline{v}_j, A\overline{v}_j\rangle = (A\overline{v}_j)^T \cdot (A\overline{v}_j) = (\overline{v}_j^T A^T)(A\overline{v}_j)\\
\Rightarrow\ &\forall j \in \overline{\{1,n\}} : ||A\overline{v}_j||^2 = \overline{v}_j^T(A^T A\overline{v}_j) = \left\langle \overline{v}_j, A^T A\overline{v}_j\right\rangle \overset{(2.30)}{=} \langle \overline{v}_j, \lambda_j\overline{v}_j\rangle\\
\Rightarrow\ &\forall j \in \overline{\{1,n\}} : ||A\overline{v}_j||^2 = \lambda_j \langle \overline{v}_j, \overline{v}_j\rangle \overset{(2.29)}{=} \lambda_j\\
\Rightarrow\ &\forall j \in \overline{\{1,n\}} : ||A\overline{v}_j||^2 = \lambda_j\\
\Rightarrow\ &\underline{\forall j \in \overline{\{1,n\}} : ||A\overline{v}_j|| = \sqrt{\lambda_j} := \sigma_j|}
\end{aligned}$$

Afirmación 4: $\{A\overline{v}_1, A\overline{v}_2, \ldots, A\overline{v}_r\}$ *es un conjunto ortogonal.*

Parte 1: $\forall k \in \overline{\{1,r\}} : A\overline{v}_k \neq \overline{0}$

En efecto, por (2.33) se tiene que:

$$\begin{array}{ccccc}
 & \sigma_1 \neq 0 & \sigma_2 \neq 0 & \ldots & \sigma_n \neq 0\\
\Rightarrow & ||A\overline{v}_1|| \neq 0 & ||A\overline{v}_2|| \neq 0 & \ldots & ||A\overline{v}_n|| \neq 0\\
\Rightarrow & A\overline{v}_1 \neq \overline{0} & A\overline{v}_2 \neq \overline{0} & \ldots & A\overline{v}_n \neq \overline{0}
\end{array}$$

Lo cual quiere decir que $\forall k \in \overline{\{1,r\}} : A\overline{v}_k \neq \overline{0}$.

Parte 2: $\forall i \neq j \in \overline{\{1, r\}} : \langle A\overline{v}_i, A\overline{v}_j \rangle = 0$

En efecto:

$$\begin{aligned}
&\forall i \neq j : \langle A\overline{v}_i, A\overline{v}_j \rangle = (A\overline{v}_i)^T \cdot (A\overline{v}_j) = (\overline{v}_i^T A^T)(A\overline{v}_j) = \overline{v}_i^T \cdot (A^T A\overline{v}_j) \\
\Rightarrow &\forall i \neq j : \langle A\overline{v}_i, A\overline{v}_j \rangle = \left\langle \overline{v}_i, A^T A\overline{v}_j \right\rangle \stackrel{(2.30)}{=} \langle \overline{v}_i, \lambda_j \overline{v}_j \rangle \stackrel{\text{Af 2.}}{=} 0
\end{aligned}$$

Por lo tanto $\{A\overline{v}_1, A\overline{v}_2, \ldots, A\overline{v}_r\}$ es un conjunto ortogonal en el espacio columna, obviamente esto se verifica por satisfacerse las afirmaciones 1 y 4 simultáneamente.

Por otro lado, al aplicar el teorema 1.1 garantiza que $\{A\overline{v}_1, A\overline{v}_2, \ldots, A\overline{v}_r\}$ es un conjunto linealmente independiente en el espacio columna.

Con todo esto sólo resta probar que cualquier elemento de $Col(A)$ es combinación lineal de los $A\overline{v}_k$ con $k \in \overline{\{1, r\}}$.

Para ello notar que por el teorema 2.8 se tiene que:

$$\forall \overline{y} \in \mathbb{R}^m : \overline{y} \in Col(A) \Rightarrow \exists \overline{x}_0 \in \mathbb{R}^n / \overline{y} = A\overline{x}_0$$

Luego, como:

$$\begin{aligned}
\overline{x}_0 \in \mathbb{R}^n &\stackrel{(2.29)}{\Rightarrow} \overline{x}_0 = \alpha_1 \overline{v}_1 + \alpha_2 \overline{v}_2 + \cdots + \alpha_n \overline{v}_n \\
&\Rightarrow A\overline{x}_0 = \alpha_1 A\overline{v}_1 + \alpha_2 A\overline{v}_2 + \cdots + \alpha_n A\overline{v}_n
\end{aligned} \tag{2.34}$$

Notar ahora lo siguiente, por (2.32) se sabe que $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$ y como por la afirmación 3 se sabe que $\forall j \in \overline{\{1, n\}} : ||A\overline{v}_j|| = \sigma_j$, se deduce entonces que:

$$\begin{aligned}
&\forall j \in \overline{\{r+1, n\}} : ||A\overline{v}_j|| = \sigma_j = 0 \\
\Rightarrow &\forall j \in \overline{\{r+1, n\}} : A\overline{v}_j = \overline{0} \\
\Rightarrow &A\overline{v}_{r+1} = A\overline{v}_{r+2} = \cdots = A\overline{v}_n = \overline{0}
\end{aligned} \tag{2.35}$$

Así, al usar (2.35) en (2.34) se obtiene que:

$$\begin{aligned}
&A\overline{x}_0 = \alpha_1 A\overline{v}_1 + \alpha_2 A\overline{v}_2 + \cdots + \alpha_r A\overline{v}_r \\
\Rightarrow &\overline{y} = \alpha_1 A\overline{v}_1 + \alpha_2 A\overline{v}_2 + \cdots + \alpha_r A\overline{v}_r
\end{aligned}$$

Obviamente, la expresión anterior está diciendo que cualquier $\overline{y}$ que está en el espacio columna de A, es combinación lineal del conjunto $\{A\overline{v}_1, A\overline{v}_2, \ldots, A\overline{v}_r\}$ y como ya se había deducido que dicho conjunto es linealmente independiente y ortogonal en el espacio columna, entonces se tiene que $\{A\overline{v}_1, A\overline{v}_2, \ldots, A\overline{v}_r\}$ es una base ortogonal de $Col(A)$. Por definición esto quiere decir que:

$$dim[Col(A)] = r \quad \Rightarrow \quad rg(A) = r$$

■

Teorema 2.15. *Descomposición de valores singulares (D.V.S)*
Sea $A_{m\times n}(\mathbb{R})$. Si A es una matriz de rango $r > 0$ entonces existe una matriz ortogonal $U_{m\times m}(\mathbb{R})$, una matriz ortogonal $V_{n\times n}(\mathbb{R})$ y una matriz $\Sigma_{m\times n}(\mathbb{R})$ de forma que $A = U\Sigma V^T$.

Donde:

$$\Sigma = \begin{bmatrix} D_{r\times r} & \vdots & \Theta_{r\times(n-r)} \\ \cdots\cdots\cdots & \cdots\cdots & \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times r} & \vdots & \Theta_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n}$$

y D es la matriz diagonal formada por los r primeros valores singulares no nulos de A $(\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0,\ \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0)$, es decir:

$$D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} =: diag(\sigma_1, \sigma_2, \ldots, \sigma_r)$$

Demostración:

Como $A_{m\times n}(\mathbb{R})$ es una matriz de rango $r > 0$ entonces por lo demostrado en el teorema 2.12 se sabe que:

$$rg(A^TA) = rg(AA^T) = rg(A) = r \tag{2.36}$$

Por otro lado, como por el teorema 2.6 se sabe que la matriz $B := A^TA \in \mathcal{M}_{n\times n}(\mathbb{R})$ es simétrica, el teorema 1.41 garantiza que B es una matriz diagonalizable y esto a su vez tiene las siguientes consecuencias.

$$\begin{cases} \text{th.1.42: } B \text{ tiene n autovalores no} \\ \quad \text{necesariamente distintos.} \\ \text{th.1.43: Existe una base ortonornal} \\ \text{de } \mathbb{R}^n \text{ formada por autovectores de } B. \end{cases} \tag{2.37}$$

Fuera de lo antes mencionado, el teorema 2.12 afirma que la matriz $B := A^T A$ es semidefinida positiva, así al unir esta condición con lo deducido en (2.37) se tiene que:

$\star$1) Existe $\mathscr{B}_0 := \{\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_n\}$ que es base ortonornal de $\mathbb{R}^n$, es decir:

$$\forall\, j,k \in \overline{\{1,n\}} : \langle \overline{v}_j, \overline{v}_k \rangle = \begin{cases} 1 & si \quad j = k \\ 0 & si \quad j \neq k \end{cases}$$

$\star$2) $\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_n$ son autovectores de $B := A^T A$.

$\star$3) Existen los escalares $\lambda_1, \lambda_2, \ldots, \lambda_n$ no necesariamente diferentes, que son autovalores de la matriz $B := A^T A$.

$\star$4) Para cada $j \in \overline{\{1,n\}}$ se tiene que $\lambda_j \geq 0$.

Asumiendo ahora que cada λ_j es un autovalor asociado a $\overline{v}_j$ y que el conjunto $\lambda_1, \lambda_2, \ldots, \lambda_n$ está ordenado en forma no creciente, se tendría que:

$$\begin{cases} \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0 \\ \forall j \in \overline{\{1,n\}} : A^T A \overline{v}_j = B\overline{v}_j := \lambda_j \overline{v}_j \end{cases} \tag{2.38}$$

Ahora bien, por lo señalado en la observación 2.3 se sabe que:

$$\begin{array}{ccccccccc} \lambda_1 & \geq & \lambda_2 & \geq & \cdots & \geq & \lambda_n & \geq & 0 \\ \Updownarrow & & \Updownarrow & & & & \Updownarrow & & \\ \sqrt{\lambda_1} & \geq & \sqrt{\lambda_2} & \geq & \cdots & \geq & \sqrt{\lambda_n} & \geq & 0 \\ \stackrel{..}{=} & & \stackrel{..}{=} & & & & \stackrel{..}{=} & & \\ \sigma_1 & \geq & \sigma_2 & \geq & \cdots & \geq & \sigma_n & \geq & 0 \end{array} \tag{2.39}$$

Por otro lado, por lo visto en la observación 1.7 se tiene que, como:

$$B_{n\times n}(\mathbb{R}) \quad \Rightarrow \quad rg(B) \leq n \quad :\Rightarrow \quad rg(A^T A) \leq n \quad \Rightarrow \quad r \leq n$$

Sin pérdida de generalidad al, asumir que $r < n$ y que los r primeros autovalores de B son no nulos, de (2.39) se obtiene que:

$$\begin{cases} \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \\ \sigma_{r+1} = \sigma_{r+1} = \cdots = \sigma_n = 0 \\ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 \\ \lambda_{r+1} = \lambda_{r+1} = \cdots = \lambda_n = 0 \end{cases} \tag{2.40}$$

Así, de (2.40) y (2.38) se concluye que:

$$A^T A\overline{v}_j := B\overline{v}_j = \begin{cases} \lambda_j & si \quad j \in \overline{\{1, r\}} \\ 0 & si \quad j \in \overline{\{r+1, n\}} \end{cases}$$

Con todo esto, es claro que, por (2.40), (2.38) y por lo indicado en ($\star 1$), ($\star 2$), ($\star 3$) y ($\star 4$) se están verificando las condiciones del teorema 2.14, por lo tanto el conjunto $\mathscr{B}_1 := \{A\overline{v}_1, A\overline{v}_2, \ldots, \overline{A}v_r\}$ es una base ortogonal del espacio $Col(A)$ y como $Col(A)$ es un subespacio de $\mathbb{R}^m$ entonces $\mathscr{B}_1$ es un conjunto ortogonal en $\mathbb{R}^m$.

Afirmación 1: $\forall j \in \overline{\{1, n\}} : \sigma_j = ||A\overline{v}_j||$

En efecto, esto se probó en la afirmación 3 del teorema 2.14.

De la afirmación 1 y (2.31) se tiene que:

$$||A\overline{v}_j|| = \begin{cases} \sigma_j > 0 & , \quad \forall j \in \overline{\{1, r\}} \\ 0 & , \quad \forall j \in \overline{\{r+1, n\}} \end{cases}$$

$$\Rightarrow \forall j \in \overline{\{r+1, n\}} : ||A\overline{v}_j|| = 0$$

$$\Rightarrow \forall j \in \overline{\{r+1, n\}} : A\overline{v}_j = \overline{0}$$

$$\Rightarrow \begin{cases} \forall j \in \overline{\{1, r\}} & : \quad ||A\overline{v}_j|| = \sigma_j > 0 \\ \forall j \in \overline{\{r+1, n\}} & : \quad A\overline{v}_j = \overline{0} \end{cases} \tag{2.41}$$

Ahora bien, como $\mathscr{B}_1 := \{A\overline{v}_1, A\overline{v}_2, \ldots, \overline{A}v_r\}$ es un conjunto ortogonal en $\mathbb{R}^m$, entonces es posible normalizar cada elemento de $\mathscr{B}_1$, al fijar:

$$\forall j \in \overline{\{1, r\}} : \overline{u}_j := \frac{1}{||A\overline{v}_j||} A\overline{v}_j \overset{(2.41)}{=} \frac{1}{\sigma_j} A\overline{v}_j \tag{2.42}$$

Con el proceso anterior, se obtiene que $\{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_r\}$ es un conjunto ortonormal en $\mathbb{R}^m$ y por tanto L.I según el teorema 1.1. Ahora, si $r = m$, el conjunto $\{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_{r=m}\}$ sería una base ortonormal de $\mathbb{R}^m$, sin embargo, si $r < m$ por el teorema de completación de bases y el teorema de ortogonalización de Gram-Schmidt, se podría extender el conjunto $\{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_r\}$ a una base ortonormal de $\mathbb{R}^m$, $\mathscr{B}_2 := \{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_r, \overline{u}_{r+1}, , \overline{u}_{r+2}, \ldots, \overline{u}_m\}$.

Luego, como la base $\mathscr{B}_2$ existe, entonces existe la matriz:

$$U := \begin{bmatrix} \overline{u}_1 & \overline{u}_2 & \cdots & \overline{u}_r & \overline{u}_{r+1} & \cdots & \overline{u}_m \end{bmatrix}$$

Asimismo, como existe la base ortonormal $\mathscr{B}_0$, entonces existe la matriz:

$$V := \begin{bmatrix} \overline{v}_1 & \overline{v}_2 & \cdots & \overline{v}_n \end{bmatrix}$$

Análogamente, como por (2.40) se sabe que que existen los valores singulares $\sigma_1, \sigma_2, \ldots, \sigma_r$ y son no nulos, entonces existe la siguiente matriz invertible.

$$D := \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} =: diag(\sigma_1, \sigma_2, \ldots, \sigma_r)$$

Finalmente, la existencia de cada una de las matrices antes mencionadas, garantiza la existencia de la siguiente matriz particionada.

$$\Sigma = \begin{bmatrix} D_{r\times r} & \vdots & \Theta_{r\times(n-r)} \\ \cdots\cdots\cdots & \cdots\cdots & \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times r} & \vdots & \Theta_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n}$$

Afirmación 2: *$U \in \mathcal{M}_{m\times m}(\mathbb{R})$ y es una matriz ortogonal.*

En efecto, como $\mathscr{B}_2$ es una base ortonormal de $\mathbb{R}^m$, se tiene que cada $\overline{u}_j$ es un vector de m coordenadas reales y como U contiene m vectores entonces se tiene que $U \in \mathcal{M}_{m\times m}(\mathbb{R})$. Además, como para cada $j \in \overline{\{1, m\}}$ se tiene que $Col_j(U) := \overline{u}_j$, entonces se tiene que las columnas de $U_{m\times m}(\mathbb{R})$ son vectores ortonormales en $\mathbb{R}^m$, lo cual quiere decir que U es una matriz ortogonal, según lo señalado en el teorema 1.8.

Afirmación 3: *$V \in \mathcal{M}_{n\times n}(\mathbb{R})$ y es una matriz ortogonal.*

En efecto, como $\mathscr{B}_0$ es una base ortonormal de $\mathbb{R}^n$, se tiene que cada $\overline{v}_j$ es un vector de n coordenadas reales y como V contiene n vectores entonces se tiene que $V \in \mathcal{M}_{n\times n}(\mathbb{R})$. Además, como para cada $j \in \overline{\{1, n\}}$ se tiene que $Col_j(V) := \overline{v}_j$, entonces se tiene que las columnas de $V_{n\times n}(\mathbb{R})$ son vectores ortonormales en $\mathbb{R}^n$, lo cual quiere decir que V es una matriz ortogonal, según lo señalado en el teorema 1.8.

Afirmación 4: $A = U\Sigma V^T$

Para ello notar que, como U y V son matrices ortogonales, entonces:

$$A = U\Sigma V^T \quad :\Leftrightarrow \quad A = U\Sigma V^{-1} \quad \Leftrightarrow \quad AV = U\Sigma$$

Por tanto, demostrar que $A = U\Sigma V^T$ equivale a demostrar que $AV = U\Sigma$.

Analizando

$$U\Sigma := \begin{bmatrix} \overline{u}_1 & \overline{u}_2 & \cdots & \overline{u}_m \end{bmatrix} \begin{bmatrix} D_{r\times r} & \vdots & \Theta_{r\times(n-r)} \\ \cdots\cdots\cdots & \cdots\cdots & \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times r} & \vdots & \Theta_{(m-r)\times(n-r)} \end{bmatrix}$$

$$:= \begin{bmatrix} \overline{u}_1 & \overline{u}_2 & \cdots & \overline{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_r & \vdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\begin{aligned}
\overset{(\text{th. } 1.44)}{=} \; & \overline{u}_1 \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \end{bmatrix} + \\
& \overline{u}_2 \begin{bmatrix} 0 & \sigma_2 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \end{bmatrix} + \cdots \\
& \overline{u}_r \begin{bmatrix} 0 & 0 & \cdots & \sigma_r & \vdots & 0 & 0 & \cdots & 0 \end{bmatrix} + \cdots \\
& \overline{u}_{r+1} \begin{bmatrix} 0 & 0 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \end{bmatrix} + \\
& \overline{u}_{r+2} \begin{bmatrix} 0 & 0 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \end{bmatrix} + \cdots \\
& \overline{u}_n \begin{bmatrix} 0 & 0 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 0 \end{bmatrix} \\
= & \begin{bmatrix} \sigma_1\overline{u}_1 & 0\overline{u}_1 & \cdots & 0\overline{u}_1 & \vdots & 0\overline{u}_1 & 0\overline{u}_1 & \cdots & 0\overline{u}_1 \end{bmatrix} + \\
& \begin{bmatrix} 0\overline{u}_2 & \sigma_2\overline{u}_2 & \cdots & 0\overline{u}_2 & \vdots & 0\overline{u}_2 & 0\overline{u}_2 & \cdots & 0\overline{u}_2 \end{bmatrix} + \cdots \\
& \begin{bmatrix} 0\overline{u}_r & 0\overline{u}_r & \cdots & \sigma_r\overline{u}_r & \vdots & 0\overline{u}_r & 0\overline{u}_r & \cdots & 0\overline{u}_r \end{bmatrix} \\
= & \begin{bmatrix} \sigma_1\overline{u}_1 & \sigma_2\overline{u}_2 & \cdots & \sigma_r\overline{u}_r & \vdots & \overline{0} & \overline{0} & \cdots & \overline{0} \end{bmatrix}
\end{aligned}$$

Por otro lado:

$$\begin{aligned}
AV &= A\left[\begin{array}{cccc} \overline{v}_1 & \overline{v}_2 & \cdots & \overline{v}_n \end{array}\right] \\
&= A\left[\begin{array}{ccccccccccccccc} \overline{v}_1 & \vdots & \overline{v}_2 & \vdots & \cdots & \vdots & \overline{v}_r & \vdots & \overline{v}_{r+1} & \vdots & \overline{v}_{r+2} & \vdots & \cdots & \vdots & \overline{v}_n \end{array}\right] \\
&= \left[\begin{array}{ccccccccccccccc} A\overline{v}_1 & \vdots & A\overline{v}_2 & \vdots & \cdots & \vdots & A\overline{v}_r & \vdots & A\overline{v}_{r+1} & \vdots & A\overline{v}_{r+2} & \vdots & \cdots & \vdots & A\overline{v}_n \end{array}\right] \\
&\overset{(2.41)}{=} \left[\begin{array}{ccccccccccccccc} A\overline{v}_1 & \vdots & A\overline{v}_2 & \vdots & \cdots & \vdots & A\overline{v}_r & \vdots & \overline{0} & \vdots & \overline{0} & \vdots & \cdots & \vdots & \overline{0} \end{array}\right] \\
&\overset{(2.42)}{=} \left[\begin{array}{ccccccccccccccc} \sigma_1\overline{v}_1 & \vdots & \sigma_2\overline{v}_2 & \vdots & \cdots & \vdots & \sigma_r\overline{v}_r & \vdots & \overline{0} & \vdots & \overline{0} & \vdots & \cdots & \vdots & \overline{0} \end{array}\right] \\
&= \left[\begin{array}{cccccccc} \sigma_1\overline{v}_1 & \sigma_2\overline{v}_2 & \cdots & \sigma_r\overline{v}_r & \vdots & \overline{0} & \overline{0} & \cdots & \overline{0} \end{array}\right] \\
&= U\Sigma
\end{aligned}$$

Por tanto queda probado que $AV = U\Sigma$ o equivalentemente $A = U\Sigma V^T$.

■

Descomposición en valores singulares (D.V.S)

Definición 2.4. *Sea $A_{m\times n}(\mathbb{R})$. La descomposición en valores singulares de A, D.V.S(A), está dada por el producto $U\Sigma V^T$, siendo U, V y Σ como en el teorema 2.15.*

Teorema 2.16. *Descomposición corta en valores singulares*

Sea $A_{m\times n}(\mathbb{R})$. Si A es una matriz de rango $r > 0$ entonces existen las matrices $P_1 \in \mathcal{M}_{m\times r}(\mathbb{R})$ y $Q_1 \in \mathcal{M}_{n\times r}(\mathbb{R})$ tales que $P_1^T P_1 = I_{r\times r} = Q_1^T Q_1$ y $A = P_1 D Q_1^T$ donde $D \in \mathcal{M}_{r\times r}(\mathbb{R})$ es una matriz diagonal con elementos diagonales positivos.

Demostración:

Como $A_{m\times n}(\mathbb{R})$ y $rg(A) = r > 0$ entonces por el teorema 2.15 se sabe que:

$$\begin{cases} \exists U_{m\times m}(\mathbb{R}) \text{ matriz ortogonal} \\ \exists V_{n\times n}(\mathbb{R}) \text{ matriz ortogonal} \\ \exists \Sigma = \left[\begin{array}{ccc} D_{r\times r} & \vdots & \Theta_{r\times(n-r)} \\ \cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times r} & \vdots & \Theta_{(m-r)\times(n-r)} \end{array}\right]_{m\times n} \end{cases} \tag{2.43}$$

De modo que

$$D = diag(\sigma_1, \sigma_2, \ldots, \sigma_r), \forall\, j \in \overline{\{1, r\}} : \sigma_j > 0 \ \wedge \ A = U\Sigma V^T$$

Ahora bien, como:

$$A = U\Sigma V^T \ \Rightarrow \ rg(A) = rg(U\Sigma V^T) \overset{\text{th.1.14}}{\leq} rg(U)$$

Luego, como por (2.43) se sabe que $U_{m\times m}(\mathbb{R})$ es una matriz ortogonal entonces por definición se tiene que U es invertible y bajo esta condición el teorema 1.20 garantiza que:

$$rg(U) = m \quad \Rightarrow \quad rg(A) \leq m \quad \Rightarrow \quad r \leq m$$

Dada la situación anterior, es posible considerar la siguiente partición:

$$U = \left[\begin{array}{ccc} P_{1(m\times r)} & \vdots & P_{2(m\times(m-r))} \end{array} \right]_{m\times m}$$

$$\Rightarrow \exists P_1 \in \mathcal{M}_{m\times r}(\mathbb{R}) \quad \wedge \quad \exists P_2 \in \mathcal{M}_{m\times(m-r)}(\mathbb{R})$$

Análogamente, como:

$$A = U\Sigma V^T \ \Rightarrow \ rg(A) = rg(U\Sigma V^T) \overset{\text{th.1.14}}{\leq} rg(V)$$

Así, como por (2.43) también se sabe que $V_{n\times n}(\mathbb{R})$ es una matriz ortogonal entonces por definición se tiene que V es invertible y bajo esta condición el teorema 1.20 garantiza que:

$$rg(V) = n \quad \Rightarrow \quad rg(A) \leq n \quad \Rightarrow \quad r \leq n$$

Por lo tanto, se puede también considerar la siguiente partición

$$V = \left[\begin{array}{ccc} Q_{1(n\times r)} & \vdots & Q_{2(n\times(n-r))} \end{array} \right]_{n\times n}$$

$$\Rightarrow \ \exists Q_1 \in \mathcal{M}_{n\times r}(\mathbb{R}) \quad \wedge \quad \exists Q_2 \in \mathcal{M}_{n\times(n-r)}(\mathbb{R})$$

Afirmación 1: $P_1^T P_1 = I_{r\times r}$

En efecto, como $U_{m\times m}(\mathbb{R})$ es una matriz ortogonal, el teorema 1.8 garantiza que sus columnas son vectores ortonormales en $\mathbb{R}^m$, luego como $P_1 \rightarrowtail U$ se infiere rápidamente que las columnas de $P_1 \in \mathcal{M}_{m\times r}(\mathbb{R})$ también son vectores ortonormales en $\mathbb{R}^m$ y bajo esta condición el teorema 1.12 garantiza que $P_1^T P_1 = I_{r\times r}$.

Afirmación 2: $Q_1^T Q_1 = I_{r\times r}$

Análogamente al caso anterior, como $V_{n\times n}(\mathbb{R})$ es una matriz ortogonal, el teorema 1.8 asegura que las columnas de V son vectores ortonormales en $\mathbb{R}^n$, luego como $Q_1 \rightarrowtail V$ entonces las columnas de $Q_1 \in \mathcal{M}_{n\times r}(\mathbb{R})$ son también vectores ortonormales en $\mathbb{R}^n$, por tanto el teorema 1.12 garantiza que $Q_1^T Q_1 = I_{r\times r}$.

Afirmación 3: $A = P_1 D Q_1^T$

En efecto:

$$
\begin{aligned}
A &= U\Sigma V^T \\
&= U\Sigma\left[\begin{array}{ccc} Q_{1(n\times r)} & \vdots & Q_{2(n\times(n-r))}\end{array}\right]^T \\
&\overset{\text{th.1.45}}{=} U\Sigma\left[\begin{array}{c} Q^T_{1(r\times n)} \\ \cdots \\ Q^T_{2((n-r)\times n)}\end{array}\right] \\
&= U\left[\begin{array}{ccc} D_{r\times r} & \vdots & \Theta_{r\times(n-r)} \\ \cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times r} & \vdots & \Theta_{(m-r)\times(n-r)}\end{array}\right]\left[\begin{array}{c} Q^T_{1(r\times n)} \\ \cdots\cdots\cdots \\ Q^T_{2((n-r)\times n)}\end{array}\right] \\
&\overset{\text{th.1.44}}{=} U\left\{\left[\begin{array}{c} D_{r\times r} \\ \cdots\cdots\cdots \\ \Theta_{(m-r)\times r}\end{array}\right]Q^T_{1(r\times n)} + \left[\begin{array}{c} \Theta_{r\times(n-r)} \\ \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times(n-r)}\end{array}\right]Q^T_{2((n-r)\times n)}\right\} \\
&= U\left\{\left[\begin{array}{c} D_{r\times r}Q^T_{1(r\times n)} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times r}Q^T_{1(r\times n)}\end{array}\right] + \left[\begin{array}{c} \Theta_{r\times(n-r)}Q^T_{2((n-r)\times n)} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times(n-r)}Q^T_{2((n-r)\times n)}\end{array}\right]\right\} \\
&= U\left\{\left[\begin{array}{c} \left(DQ_1^T\right)_{r\times n} \\ \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times n}\end{array}\right] + \left[\begin{array}{c} \Theta_{r\times n} \\ \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times n}\end{array}\right]\right\} \\
&= U\left[\begin{array}{c} \left(DQ_1^T\right)_{r\times n} \\ \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times n}\end{array}\right] \\
&= \left[\begin{array}{ccc} P_{1(m\times r)} & \vdots & P_{2(m\times(m-r))}\end{array}\right]\left[\begin{array}{c} \left(DQ_1^T\right)_{r\times n} \\ \cdots\cdots\cdots\cdots \\ \Theta_{(m-r)\times n}\end{array}\right]
\end{aligned}
$$

Así, al aplicar nuevamente el teorema 1.44 en la última igualdad, se obtiene que:

$$\begin{aligned} A &= P_{1(m\times r)}\left(DQ_1^T\right)_{r\times n} + P_{2(m\times(m-r))}\Theta_{(m-r)\times n} \\ &= \left(P_1DQ_1^T\right)_{m\times n} + \Theta_{m\times n} \\ &:= P_1DQ_1^T \end{aligned}$$

■

Teorema 2.17.
Sean $A, B \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si $A^TA = B^TB$ entonces existe una matriz ortogonal $Q_0 \in \mathcal{M}_{m\times m}(\mathbb{R})$ de modo que $B = Q_0A$.

Demostración:

Por el teorema D.V.S(A) se tiene que:
$\exists(Q_A)_{m\times m}(\mathbb{R})$: Matriz ortogonal.
$\exists(V_A)_{n\times n}(\mathbb{R})$: Matriz ortogonal.
$\exists(\Sigma_A)_{m\times m}(\mathbb{R})$: Matriz que contiene los valores singulares no nulos de A.

de modo que:

$$A = Q_A\Sigma_AV_A^T$$

Análogamente, por el teorema D.V.S(B) se tiene que:
$\exists(Q_B)_{m\times m}(\mathbb{R})$: Matriz ortogonal.
$\exists(V_B)_{n\times n}(\mathbb{R})$: Matriz ortogonal.
$\exists(\Sigma_B)_{m\times m}(\mathbb{R})$: Matriz que contiene los valores singulares no nulos de B.

de modo que:

$$B = Q_B\Sigma_BV_B^T$$

Ahora bien, por el teorema 2.12 se sabe que:

$$rg(A) = rg(A^TA) \quad \wedge \quad rg(B) = rg(B^TB)$$

Pero como por hipótesis se sabe que:

$$A^TA = B^TB \quad \Rightarrow \quad rg(A^TA) = rg(B^TB) \quad \Rightarrow \quad rg(A) = rg(B) := r$$

Asimismo, como $A^TA = B^TB$ por definición se deduce que A y B tienen los mismos valores singulares y como $rg(A) = rg(B) := r$ entonces se tiene que:

$$\Sigma_A = \Sigma_B := \Sigma_0$$

El que $A^TA = B^TB$ también implica que $P_\lambda(A^TA) = P_\lambda(B^TB)$, es decir A^TA y B^TB tienen los mismos autovalores, consecuentemente A^TA y B^TB tienen los mismos autovectores. Luego, como en la prueba del teorema 2.15 se ve que V_A es la matriz formada por una base ortnornal de autovectores de A^TA, por lo antes mencionado se deduce que:

$$V_A = V_B := V_0$$

Por tanto, es claro que:

$$\begin{aligned}
& A = Q_A\Sigma_0V_0^T \quad \wedge \quad B = Q_B\Sigma_0V_0^T \\
\Rightarrow\ & \Sigma_0V_0^T = Q_A^{-1}A \quad \wedge \quad B = Q_B\Sigma_0V_0^T \\
:\Rightarrow\ & \Sigma_0V_0^T = Q_A^TA \quad \wedge \quad B = Q_B\Sigma_0V_0^T \\
\Rightarrow\ & B = Q_BQ_A^TA \\
\Rightarrow\ & \exists\, Q_0 := Q_BQ_A^T \in \mathcal{M}_{m\times m}(\mathbb{R}) \,/\, B = Q_0A
\end{aligned}$$

Afirmación: $Q_0 := Q_BQ_A^T \in \mathcal{M}_{m\times m}(\mathbb{R})$ *es una matriz ortogonal*

En efecto, como $Q_A, Q_B \in \mathcal{M}_{m\times m}(\mathbb{R})$ son ambas matrices ortogonales, el teorema 1.11 garantiza que el producto de ellas también es una matriz ortogonal, por tanto $Q_0 := Q_BQ_A^T$ es una matriz ortogonal.

Se concluye entonces que existe una matriz ortogonal $Q_0 :\in \mathcal{M}_{m\times m}(\mathbb{R})$ de modo que $B = Q_0A$.

■

Teorema 2.18.
Sean $X, Y \in \mathcal{M}_{m\times n}(\mathbb{R})$ y sean $B, C \in \mathcal{M}_{m\times m}(\mathbb{R})$ un par de marices simétricas definidas positivas.
Si $X^TB^{-1}X = Y^TC^{-1}Y$ entonces existe una matriz invertible $A_{m\times m}(\mathbb{R})$ con la cual se verifica que:

$$Y = AX \quad \wedge \quad C = ABA^T$$

Demostración:

Como $B_{m\times m}(\mathbb{R})$ y $C_{m\times m}(\mathbb{R})$ son matrices simétricas y definidas positivas, el teorema 2.9 garantiza lo siguiente:

$$\begin{cases} \exists B^{-1}, C^{-1} \in \mathcal{M}_{m\times m}(\mathbb{R}) \\ \exists B^{1/2}, C^{1/2} \in \mathcal{M}_{m\times m}(\mathbb{R}) \text{ simétricas} \\ B = B^{1/2}B^{1/2} \quad \wedge \quad C = C^{1/2}C^{1/2} \end{cases} \tag{2.44}$$

Además se verifica que:

$$\begin{cases} \exists B^{-1/2}, C^{-1/2} \in \mathcal{M}_{m\times m}(\mathbb{R}) \ \textit{simétricas} \\ B^{-1} = B^{-1/2}B^{-1/2} \quad \wedge \quad C^{-1} = C^{-1/2}C^{-1/2} \\ B^{-1/2}B^{1/2} = I_{m\times m} = B^{1/2}B^{-1/2} \\ C^{-1/2}C^{1/2} = I_{m\times m} = C^{1/2}C^{-1/2} \end{cases} \tag{2.45}$$

Así, dada la existencia $B^{-1/2}$ y $C^{-1/2}$ se tiene que las siguientes matrices existen y están bien definidas

$$X_1 := B^{-1/2}X \quad , \quad Y_1 := C^{-1/2}Y \tag{2.46}$$

Afirmación 1: $X_1^T X_1 = Y_1^T Y_1$

En efecto:

$$\begin{aligned} &X_1^T X_1 := (B^{-1/2}X)^T(B^{-1/2}X) \overset{th.}{=} X^T(B^{-1/2})^T B^{-1/2}X \overset{(2.44)}{=} X^T B^{-1/2}B^{-1/2}X \\ \Rightarrow &X_1^T X_1 \overset{(2.45)}{=} X^T B^{-1}X = Y^T C^{-1}Y \overset{(2.45)}{=} Y^T(C^{-1/2}C^{-1/2})Y \\ \Rightarrow &X_1^T X_1 \overset{(2.45)}{=} Y^T(C^{-1/2})^T(C^{-1/2}Y) \overset{th.}{=} (C^{-1/2}Y)^T(C^{-1/2}Y) := Y_1^T Y_1 \end{aligned}$$

Afirmación 2: *Existe una matriz ortogonal Q_0 tal que* $Y_1 = Q_0X_1$

En efecto, por la forma en que se definió X_1, Y_1 se tiene $X_1, Y_1 \in M_{m\times n}(\mathbb{R})$, además por la afirmación 1 se sabe que $X_1^T X_1 = Y_1^T Y_1$ entonces bajo estas condiciones el teorema 2.17 garantiza que existe una matriz ortogonal $Q_0 \in \mathcal{M}_{m\times m}(\mathbb{R})$ de modo que $Y_1 = Q_0X_1$.

Así, de lo antes mencionado y de (2.46) se tiene que:

$$\begin{aligned} C^{-1/2}Y = Q_0B^{-1/2}X &\Rightarrow C^{1/2}C^{-1/2}Y = C^{1/2}Q_0B^{-1/2}X \\ &\overset{(2.45)}{\Rightarrow} Y = C^{1/2}Q_0B^{-1/2}X \\ &\Rightarrow \exists A := C^{1/2}Q_0B^{-1/2} / Y = AX \end{aligned} \tag{2.47}$$

Afirmación 3: $A \in \mathcal{M}_{m\times m}(\mathbb{R})$ *y además es invertible.*

En efecto, como $C^{1/2}$, Q_0 y $B^{-1/2}$ son elementos del espacio $\mathcal{M}_{m\times m}(\mathbb{R})$, se tiene que $C^{1/2}Q_0B^{-1/2} \in \mathcal{M}_{m\times m}(\mathbb{R})$, es decir $A \in \mathcal{M}_{m\times m}(\mathbb{R})$.

Por otro lado, por (2.44) y (2.45) se sabe que $B^{-1/2}$ y $C^{1/2}$ son matrices invertibles, ahora como Q_0 es invertible por ser una matriz ortogonal entonces el producto $C^{1/2}Q_0B^{-1/2}$ es una matriz invertible, es decir A es una matriz invertible.

Afirmación 4: $C = ABA^T$

En efecto:

$$\begin{aligned}
ABA^T &:= (C^{1/2}Q_0B^{-1/2})B(C^{1/2}Q_0B^{-1/2})^T \\
&\overset{(2.44)}{=} (C^{1/2}Q_0\,\underline{B^{-1/2})B^{1/2}}\,B^{1/2}(C^{1/2}Q_0B^{-1/2})^T \\
&= (C^{1/2}Q_0)B^{1/2}(C^{1/2}Q_0B^{-1/2})^T \\
&\overset{th.}{=} (C^{1/2}Q_0)B^{1/2}(B^{-1/2})^TQ_0^T(C^{1/2})^T \\
&\overset{(2.45)}{=} (C^{1/2}Q_0)\,\underline{B^{1/2}B^{-1/2}}\,Q_0^T(C^{1/2})^T \\
&\overset{(2.45)}{=} (C^{1/2}\,\underline{Q_0)Q_0^T}\,(C^{1/2})^T \\
&:= C^{1/2}(C^{1/2})^T \\
&\overset{(2.44)}{=} C^{1/2}C^{1/2} \\
&\overset{(2.44)}{=} C
\end{aligned}$$

Así, de (2.47) de la afirmación 3 y la afirmación 4 se sigue que:

$$\exists A \in \mathcal{M}_{m\times m}(\mathbb{R}) \,/\, \exists A^{-1}, Y = AX \quad \wedge \quad C = ABA^T$$

■

2.4. Inversa de Moore-Penrose y propiedades

Inversa generalizada de Moore-Penrose

Definición 2.5. *Sea $A_{m\times n}(\mathbb{R})$ y sea $X_{n\times m}(\mathbb{R})$. La matriz X es la inversa generalizada de Moore-Penrose de la matriz A, si y sólo si, se verifican las siguientes condiciones.*

1. $AXA = A$.
2. $XAX = X$.
3. XA *es una matriz simétrica, es decir* $(XA)^T = XA$.
4. AX *es una matriz simétrica, es decir* $(AX)^T = AX$.

Observación 2.4.

▷ *Las expresiones "inversa generalizada de Moore-Pernose" e "inversa de Moore-Penrose", son sinónimas.*

▷ *La matriz D que contiene los valores singulares no nulos A, indicada en los teoremas de descomposición 2.15 y 2.16 será denotada por Δ.*

▷ *Cuando una matriz X satisface las condiciones anteriores, se denota por $A^\dagger$.*

▷ *Del punto anterior y de la definición se obtiene que:*

$$X = A^\dagger :\Leftrightarrow AXA = A \quad , \quad XAX = X \quad , \quad (XA)^T = XA \quad , \quad (AX)^T = AX$$

Teorema 2.19.

La inversa de Moore-Penrose de la matriz $\Theta_{m\times n}(\mathbb{R})$, es la matriz $\Theta_{n\times m}(\mathbb{R})$.

Demostración:

Dada la matriz $A := \Theta_{m\times n}(\mathbb{R})$ y la matriz $X := \Theta_{n\times m}(\mathbb{R})$ se tiene que:

- $AXA := \Theta_{m\times n}(\Theta_{n\times m})\Theta_{m\times n} = \Theta_{m\times n} := A$
- $XAX := \Theta_{n\times m}(\Theta_{m\times n})\Theta_{n\times m} = \Theta_{n\times m} := X$
- $(XA)^T := (\Theta_{n\times m}\Theta_{m\times n})^T = (\Theta_{n\times n})^T = \Theta_{n\times n} = \Theta_{n\times m}\Theta_{m\times n} = XA$
- $(AX)^T := (\Theta_{m\times n}\Theta_{n\times m})^T = (\Theta_{m\times m})^T = \Theta_{m\times m} = \Theta_{m\times n}\Theta_{n\times m} = AX$

Por lo tanto $X = A^\dagger$, es decir $\Theta_{n\times m} = (\Theta_{m\times n})^\dagger$, es decir $(\Theta_{m\times n})^\dagger = \Theta_{n\times m}$.

■

Teorema 2.20.

Existe la inversa de Moore-Penrose de cualquier matriz $A_{m\times n}(\mathbb{R})$.

Demostración:

En efecto, si $A = \Theta_{m\times n}(\mathbb{R})$ entonces $A^\dagger = \Theta_{n\times m}$, según lo que se demostró en el teorema anterior.

Ahora bien, si:

$$\begin{aligned} A \neq \Theta_{m\times n} \quad \Rightarrow \quad \dim[Col(A)] > 0 \quad &\Rightarrow \quad \exists r \in \mathbb{N} \,/\, \dim[Col(A)] = r \\ &:\Rightarrow \quad \exists r \in \mathbb{N} \,/\, rg(A) = r > 0 \end{aligned}$$

Bajo estas condiciones, el teorema 2.16, garantiza que:

$$\begin{cases} \exists P_1 \in M_{m\times r}(R) \quad , \quad \exists Q_1 \in M_{n\times r}(R) \\ P_1^T P_1 = I_{r\times r} = Q_1^T Q_1 \\ A = P_1 \Delta Q_1^T \end{cases} \tag{2.48}$$

Donde $\Delta \in \mathcal{M}_{r\times r}(\mathbb{R})$ es una matriz diagonal, cuyos elementos diagonales son positivos, lo cual implica que $det(\Delta) \neq 0$, es decir $\exists \Delta^{-1}$, por lo tanto:

$$\exists X := Q_1 \Delta^{-1} P_1^T$$

Afirmación: $X = A^{\dagger}$

En efecto:

- $$\begin{aligned} AXA &\overset{(2.48)}{=} (P_1\Delta Q_1^T)X(P_1\Delta Q_1^T) \\ &:= (P_1\Delta \underline{Q_1^T)(Q_1} \Delta^{-1} \underline{P_1^T)(P_1} \Delta Q_1^T) \\ &\overset{(2.48)}{=} (P_1\Delta)\,\underline{\Delta^{-1}(\Delta}\, Q_1^T) \\ &:= P_1 \Delta Q_1^T \\ &\overset{(2.48)}{=} A \end{aligned}$$

- $$\begin{aligned} XAX &:= (Q_1\Delta^{-1}P_1^T)A(Q_1\Delta^{-1}P_1^T) \\ &\overset{(2.48)}{=} (Q_1\Delta^{-1} \underline{P_1^T)(P_1} \Delta \underline{Q_1^T)(Q_1} \Delta^{-1}P_1^T) \\ &\overset{(2.48)}{=} (Q_1\Delta^{-1})\,\underline{\Delta(\Delta^{-1}}\, P_1^T) \\ &:= Q_1\Delta^{-1}P_1^T \\ &:= X \end{aligned}$$

- $$\begin{aligned} (XA)^T &= [(Q_1\Delta^{-1} \underline{P_1^T)(P_1} \Delta Q_1^T)]^T \\ &\overset{(2.48)}{=} [(Q_1 \underline{\Delta^{-1})(\Delta} Q_1^T)]^T \\ &:= (Q_1Q_1^T)^T \\ &\overset{\text{th.2.6}}{=} Q_1Q_1^T \\ &= (Q_1\Delta^{-1}P_1^T)(P_1\Delta Q_1^T) \\ &= XA \end{aligned}$$

- $$\begin{aligned} (AX)^T &= [(P_1\Delta \underline{Q_1^T)(Q_1} \Delta^{-1}P_1^T)]^T \overset{(2.48)}{=} [(P_1 \underline{\Delta)(\Delta^{-1}} P_1^T)]^T := (P_1P_1^T)^T \\ &\overset{\text{th.2.6}}{=} P_1P_1^T = (P_1\Delta Q_1^T)(Q_1\Delta^{-1}P_1^T) = AX \end{aligned}$$

Por lo tanto $A^\dagger = X$ o equivalentemente $X = A^\dagger$.

De este modo, queda probado que:

$$\begin{cases} A = \Theta_{m\times n} \Rightarrow \exists X := \Theta_{n\times m}/A^\dagger = X \\ A \neq \Theta_{m\times n} \Rightarrow \exists X := Q_1\Delta^{-1}P_1^T/A^\dagger = X \end{cases}$$

Por lo tanto existe la inversa de Moore-Penrose de cualquiera matriz $A_{m\times n}(\mathbb{R})$.

■

Teorema 2.21.

Existe una única inversa de Moore-Penrose de cualquier matriz $A_{m\times n}(\mathbb{R})$.

Demostración:

La existencia se tiene garantizada por el teorema anterior.

Ahora, si $A_1^\dagger, A_1^\dagger \in \mathcal{M}_{n\times m}(\mathbb{R})$ son dos inversas de Moore-Penrose distintas, entonces para cualquier $j \in \{1, 2\}$, se tiene que:

$$AA_j^\dagger A = A \quad , \quad A_j^\dagger AA_j^\dagger = A_j^\dagger \quad , \quad (A_j^\dagger A)^T = A_j^\dagger A \quad , \quad (AA_j^\dagger)^T = AA_j^\dagger$$

En base a estas igualdades matriciales se tiene que:

$$A_1^\dagger = A_1^\dagger\, \underline{A}\, A_1^\dagger = A_1^\dagger(\underline{AA_2^\dagger}\, \underline{A})A_1^\dagger = A_1^\dagger(AA_2^\dagger)^T(AA_1^\dagger)^T \overset{th.}{=} A_1^\dagger[(A_2^\dagger)^T\, \underline{A^T][(A_1^\dagger)^T A^T}]$$

$$\Rightarrow A_1^\dagger \overset{th.}{=} A_1^\dagger(A_2^\dagger)^T(AA_1^\dagger A)^T = A_1^\dagger\, \underline{(A_2^\dagger)^T A^T} \overset{th.}{=} A_1^\dagger(AA_2^\dagger)^T = A_1^\dagger\, \underline{A}\, A_2^\dagger$$

$$\Rightarrow A_1^\dagger = \underline{A_1^\dagger(A}\ \ \underline{A_2^\dagger A})A_2^\dagger = (A_1^\dagger A)^T(A_2^\dagger A)^T A_2^\dagger \overset{th.}{=} [\underline{A^T(A_1^\dagger)^T][A^T}\, (A_2^\dagger)^T]A_2^\dagger$$

$$\Rightarrow A_1^\dagger \overset{th.}{=} (AA_1^\dagger A)^T(A_2^\dagger)^T A_2^\dagger = \underline{A^T(A_2^\dagger)^T}\, A_2^\dagger \overset{th.}{=} (A_2^\dagger A)^T A_2^\dagger = A_2^\dagger AA_2^\dagger = A_2^\dagger$$

$$\Rightarrow A_1^\dagger = A_2^\dagger \quad (\Rightarrow\Leftarrow)$$

La igualdad anterior representa una contradicción porque desde un inicio se asumió que $A_1^\dagger \neq A_2^\dagger$.

■

Teorema 2.22.

Dada una matriz $A_{m\times n}(\mathbb{R})$ y una matriz $B_{n\times m}(\mathbb{R})$, se cumple que:

$$A = B^\dagger \quad \Leftrightarrow \quad B = A^\dagger$$

Demostración:

Por definición se tiene que:

$$X = R^\dagger :\Leftrightarrow RXR = R\ ,\ XRX = X\ ,\ (XR)^T = XR\ ,\ (RX)^T = RX \tag{2.49}$$

De (2.49) se sigue que, si:

$$\begin{aligned} A = B^\dagger &\Rightarrow BAB = B\ ,\ ABA = X\ ,\ (AB)^T = AB\ ,\ (BA)^T = BA \\ &\Rightarrow ABA = A\ ,\ BAB = B\ ,\ (BA)^T = BA\ ,\ (AB)^T = AB \\ &\Rightarrow B = A^\dagger \end{aligned}$$

Asimismo, si

$$\begin{aligned} B = A^\dagger &\Rightarrow ABA = A\ ,\ BAB = B\ ,\ (BA)^T = BA\ ,\ (AB)^T = AB \\ &\Rightarrow BAB = B\ ,\ ABA = A\ ,\ (AB)^T = AB\ ,\ (BA)^T = BA \\ &\Rightarrow A = B^\dagger \end{aligned}$$

De este modo, queda probado que $A = B^\dagger$ si y sólo si $B = A^\dagger$.

■

Teorema 2.23.

Dada una matriz $A_{m\times n}(\mathbb{R})$ se cumple que:

1. $(A^\dagger)^\dagger = A$.
2. $\forall\lambda \in R : (\lambda A)^\dagger = \lambda^\dagger A^\dagger$, *donde* $\lambda^\dagger = \begin{cases} \dfrac{1}{\lambda} & , \quad \lambda \neq 0 \\ 0 & , \quad \lambda = 0 \end{cases}$.
3. $(A^T)^\dagger = (A^\dagger)^T$.
4. $(AA^T)^\dagger = (A^T)^\dagger A^\dagger$.
5. $(A^TA)^\dagger = A^\dagger(A^T)^\dagger$.
6. $A^\dagger = (A^TA)^\dagger A^T = A^T(AA^T)^\dagger$.
7. $A^T = A^TAA^\dagger = A^\dagger AA^T$.

Demostración:

Por definición se tiene que:

$$X = R^\dagger :\Leftrightarrow RXR = R\ ,\ XRX = X\ ,\ (XR)^T = XR\ ,\ (RX)^T = RX \tag{2.50}$$

(1)

Dada la matriz $A_{m\times n}(\mathbb{R})$, por el teorema 2.21 se tiene que

$$
\begin{aligned}
&\exists! X \in \mathcal{M}_{n\times m}(\mathbb{R}) \,/\, X = A^{\dagger} \\
&\overset{(2.50)}{\Rightarrow} AXA = A \ , \ XAX = X \ , \ (XA)^T = XA \ , \ (AX)^T = AX \\
&\Rightarrow XAX = X \ , \ AXA = A \ , \ (AX)^T = AX \ , \ (XA)^T = XA \\
&\Rightarrow A^{\dagger}AA^{\dagger} = A^{\dagger} \ , \ AA^{\dagger}A = A \ , \ (AA^{\dagger})^T = AA^{\dagger} \ , \ (A^{\dagger}A)^T = A^{\dagger}A \\
&\overset{(2.50)}{\Rightarrow} A = (A^{\dagger})^{\dagger}
\end{aligned}
$$

(2)

Si $\lambda = 0$, por hipótesis se tiene que $\lambda^{\dagger} = 0$, lo cual implica que:

$$
\begin{cases}
(\lambda A)^{\dagger} = (0A)^{\dagger} = \Theta^{\dagger} \overset{(\text{th.}2.19)}{=} \Theta \\
\lambda^{\dagger}A^{\dagger} = 0A^{\dagger} = \Theta
\end{cases}
$$

Por lo tanto $(\lambda A)^{\dagger} = \lambda^{\dagger}A^{\dagger}$.

Por otro lado, si $\lambda \neq 0$ por hipótesis se tiene que $\lambda^{\dagger} := \dfrac{1}{\lambda}$, luego, como por el teorema de existencia y unicidad 2.21 se tiene que existe la matriz $A^{\dagger}$, se deduce que:

$$
\begin{aligned}
&\exists X := \lambda^{\dagger}A^{\dagger} \quad \wedge \quad \exists R := \lambda A \\
\Leftrightarrow \ &\exists X := \frac{1}{\lambda}A^{\dagger} \quad \wedge \quad \exists R := \lambda A
\end{aligned}
$$

Ahora:

- $RXR = (\lambda A)\left(\dfrac{1}{\lambda}A^{\dagger}\right)(\lambda A) = \lambda AA^{\dagger}A = \lambda A := R$

- $XRX = \left(\dfrac{1}{\lambda}A^{\dagger}\right)(\lambda A)\left(\dfrac{1}{\lambda}A^{\dagger}\right) = \dfrac{1}{\lambda}A^{\dagger}AA^{\dagger} = \dfrac{1}{\lambda}A^{\dagger} := X$

- $(XR)^T = \left[\left(\dfrac{1}{\lambda}A^{\dagger}\right)(\lambda A^{\dagger})\right]^T = (A^{\dagger}A)^T := A^{\dagger}A = \left(\dfrac{1}{\lambda}A^{\dagger}\right)(\lambda A^{\dagger}) := XR$

- $(RX)^T = \left[(\lambda A)\left(\dfrac{1}{\lambda}A^{\dagger}\right)\right]^T = (AA^{\dagger})^T := AA^{\dagger} = (\lambda A)\left(\dfrac{1}{\lambda}A^{\dagger}\right) := RX$

Así, por (2.50) se sigue que $X = R^{\dagger}$ o en forma equivalente $R^{\dagger} = X$, es decir $(\lambda A)^{\dagger} = \lambda^{\dagger}A^{\dagger}$. para cualquier $\lambda \neq 0$, pero cuando $\lambda = 0$, se vio que dicha igualdad también se satisface, es decir $(\lambda A)^{\dagger} = \lambda^{\dagger}A^{\dagger}$ para cualquier $\lambda \in \mathbb{R}$.

(3)
Tomando $X = (A^\dagger)^T$ y $R = A^T$ se tiene que:

- $RXR = A^T(A^\dagger)^T A^T \overset{th.}{=} (AA^\dagger A)^T := A^T := R$
- $XRX = (A^\dagger)^T A^T (A^\dagger)^T \overset{th.}{=} (A^\dagger A A^\dagger)^T := (A^\dagger)^T := X$
- $(XR)^T = [(A^\dagger)^T A^T]^T \overset{th.}{=} [(AA^\dagger)^T]^T \overset{th.}{=} AA^\dagger := (AA^\dagger)^T \overset{th.}{=} (A^\dagger)^T A^T := XR$
- $(RX)^T = [A^T(A^\dagger)^T]^T \overset{th.}{=} [(A^\dagger A)^T]^T \overset{th.}{=} A^\dagger A := (A^\dagger A)^T \overset{th.}{=} A^T(A^\dagger)^T := RX$

Se concluye por tanto de (2.50) que $R^\dagger = X$, es decir $(A^T)^\dagger = (A^\dagger)^T$.

(4)
Tomando $X := (A^T)^\dagger A^\dagger$ y $R = AA^T$, se tiene que:

- $RXR := AA^T[(A^T)^\dagger A^\dagger]AA^T \overset{(3)}{=} AA^T[(A^\dagger)^T \underline{A^\dagger]A}\, A^T := AA^T(A^\dagger)^T(A^\dagger A)^T A^T$

$\Rightarrow$ $RXR \overset{th.}{=} AA^T(A^\dagger)^T \underline{[A^T(A^\dagger)^T]}A^T \overset{th.}{=} AA^T(A^\dagger)^T(AA^\dagger A)^T := A\,\underline{A^T(A^\dagger)^T}A^T$

$\Rightarrow$ $RXR \overset{th.}{=} A(AA^\dagger A)^T := AA^T := R$

- $XRX := [(A^T)^\dagger A^\dagger]AA^T[(A^T)^\dagger A^\dagger] \overset{(3)}{=} [(A^T)^\dagger A^\dagger]A\,\underline{A^T[(A^\dagger)^T}\, A^\dagger]$

$\Rightarrow$ $XRX \overset{th.}{=} [(A^T)^\dagger A^\dagger]A(A^\dagger A)^T A^\dagger := [(A^T)^\dagger A^\dagger]A\,\underline{(A^\dagger A)A^\dagger} := [(A^T)^\dagger\, \underline{A^\dagger]AA^\dagger}$

$\Rightarrow$ $XRX = (A^T)^\dagger A^\dagger := X$

Ahora bien, por la forma en que se ha definido X y R, se tiene que:

$$XR := [(A^T)^\dagger\, \underline{A^\dagger]A}\, A^T := (A^T)^\dagger\, \underline{(A^\dagger A)^T A^T} \overset{th.}{=} (A^T)^\dagger(AA^\dagger A)^T := (A^T)^\dagger A^T$$

$$\Rightarrow XR \overset{(3)}{=} (A^\dagger)^T A^T \overset{th.}{=} (AA^\dagger)^T := AA^\dagger \tag{2.51}$$

Ahora bien:

- $(XR)^T := \{[(A^T)^\dagger A^\dagger]AA^T\}^T \overset{th.}{=} (AA^T)^T[(A^T)^\dagger A^\dagger]^T \overset{(2.6)}{=} AA^T[(A^T)^\dagger A^\dagger]^T$

$\Rightarrow$ $(XR)^T \overset{th.}{=} AA^T(A^\dagger)^T[(A^T)^\dagger]^T \overset{(3)}{=} AA^T(A^\dagger)^T[(A^\dagger)^T]^T \overset{th.}{=} A\,\underline{A^T(A^\dagger)^T}\, A^\dagger$

De este modo se obtiene que:

$$\gtrdot \quad (XR)^T \overset{th.}{=} A(A^\dagger A)^T A^\dagger := A\,\underline{(A^\dagger A)}A^\dagger := AA^\dagger \overset{(2.51)}{=} XR$$

Asimismo, se observa que:

$$RX := AA^T[(A^T)^\dagger A^\dagger] \overset{(3)}{=} A\,\underline{A^T[(A^\dagger)^T}\,A^\dagger] \overset{th.}{=} A(A^\dagger A)^T A^\dagger := \underline{A(A^\dagger A)}\,A^\dagger$$

$$\Rightarrow RX = AA^\dagger \tag{2.52}$$

Con esto en mente, se obtiene lo siguiente.

$$\gtrdot \quad (RX)^T := \{AA^T[(A^T)^\dagger A^\dagger]\}^T \overset{th.}{=} [(A^T)^\dagger A^\dagger]^T (AA^T)^T \overset{th.}{=} (A^\dagger)^T[(A^T)^\dagger]^T(AA^T)$$

$$\Rightarrow (RX)^T \overset{(3)}{=} (A^\dagger)^T[(A^\dagger)^T]^T(AA^T) \overset{th.}{=} (A^\dagger)^T\,\underline{A^\dagger(A}\,A^T) := (A^\dagger)^T(A^\dagger A)^T A^T$$

$$\Rightarrow (RX)^T \overset{th.}{=} (A^\dagger)^T\,[\underline{A^T(A^\dagger)^T}]A^T \overset{th.}{=} (A^\dagger)^T(AA^\dagger A)^T := (A^\dagger)^T A^T \overset{th.}{=} (AA^\dagger)^T$$

$$\Rightarrow (RX)^T = AA^\dagger \overset{(2.52)}{=} RX$$

Por tanto de (2.50) se concluye que $R^\dagger = X$, es decir $(AA^T)^\dagger = (A^T)^\dagger A^\dagger$.

(5)

En ítem anterior se ha probado que $(BB^T)^\dagger = (B^T)^\dagger B^\dagger$, así, al considerar $B = A^T$ se tendrá que:

$$[A^T(A^T)^T]^\dagger = [(A^T)^T]^\dagger(A^T)^\dagger \quad \Rightarrow \quad (A^TA)^\dagger = A^\dagger(A^T)^\dagger$$

(6)

Para esto notar que:

$$\gtrdot \quad A^\dagger := \underline{A^\dagger A}\,A^\dagger := (A^\dagger A)^T A^\dagger \overset{th.}{=} A^T(A^\dagger)^T A^\dagger \overset{(3)}{=} A^T(A^T)^\dagger A^\dagger \overset{(4)}{=} A^T(AA^T)^\dagger$$

$$\gtrdot \quad A^\dagger := A^\dagger\,\underline{AA^\dagger} := A^\dagger(AA^\dagger)^T \overset{th.}{=} A^\dagger(A^\dagger)^T A^T \overset{(3)}{=} A^\dagger(A^T)^\dagger A^T \overset{(5)}{=} (A^TA)^\dagger A^T$$

$$\Rightarrow \quad A^\dagger = A^T(AA^T)^\dagger = (A^TA)^\dagger A^T$$

(7)

Al igual que el caso anterior, notar que:

$$\gtrdot \quad A^T := (AA^\dagger A)^T \overset{th.}{=} A^T\,\underline{(A^\dagger)^T}\,A^T \overset{th.}{=} A^T(AA^\dagger)^T := A^TAA^\dagger$$

Por otro lado:

$$\gtrdot \quad A^T := (AA^{\dagger}A)^T \overset{th.}{=} \underline{A^T(A^{\dagger})^T}\, A^T \overset{th.}{=} (A^{\dagger}A)^T A^T := A^{\dagger}AA^T$$

$$\Rightarrow \quad A^T = A^T AA^{\dagger} = (A^{\dagger}A)^T A^T$$

■

Teorema 2.24.

Dada una matriz $A_{m\times n}(\mathbb{R})$ se cumple que:

1. Si A es una matriz cuadrada e invertible, entonces $A^{\dagger} = A^{-1}$.

2. Si $rg(A) = n$ entonces $A^{\dagger}A = I_{n\times n}$.

3. Si $rg(A) = m$ entonces $AA^{\dagger} = I_{m\times m}$.

Demostración:

Por definición se tiene que:

$$X = R^{\dagger} :\Leftrightarrow RXR = R\ ,\ XRX = X\ ,\ (XR)^T = XR\ ,\ (RX)^T = RX \tag{2.53}$$

(1)

Considerando $X := A^{-1}$ y $R := A$, que puede hacerse porque al ser A una matriz cuadrada e invertible existe A^{-1}, puede notarse que:

$$\gtrdot \quad RXR := \underline{AA^{-1}}\,A := I_nA = A := R$$

$$\gtrdot \quad XRX := \underline{A^{-1}A}\,A^{-1} := I_nA^{-1} = A^{-1} := X$$

$$\gtrdot \quad (XR)^T := (A^{-1}A)^T = I_n^T = I_n = A^{-1}A := XR$$

$$\gtrdot \quad (RX)^T := (AA^{-1})^T = I_n^T = I_n = AA^{-1} := RX$$

Por tanto, de (2.53) se obtiene que $A^{\dagger} = A^{-1}$.

(2)

Si se cumple que:

$$rg(A) = n \overset{\text{th.2.12}}{\Rightarrow} rg(A) = rg(A^TA) = rg(AA^T) = n$$

Por otro lado, como:

$$A \in \mathcal{M}_{m\times n}(\mathbb{R}) :\Rightarrow A^T \in \mathcal{M}_{n\times m}(\mathbb{R}) \Rightarrow A^TA \in \mathcal{M}_{n\times n}(\mathbb{R})$$
$$\Rightarrow \quad A^TA \in \mathcal{M}_{n\times n}(\mathbb{R}) \quad \wedge \quad rg(A^TA) = n$$

Luego, bajo las condiciones anteriores el teorema 1.20 garantiza que:

$$\exists(A^TA)^{-1} \overset{(1)}{\Rightarrow} (A^TA)^{\dagger} = (A^TA)^{-1}$$

Ahora bien, por el teorema 2.23 se sabe que:

$$\begin{aligned} & A^{\dagger} = (A^TA)^{\dagger}A^T = A^T(AA^T)^{\dagger} \\ \Rightarrow\ & A^{\dagger} = (A^TA)^{\dagger}A^T \\ \Rightarrow\ & A^{\dagger} = (A^TA)^{-1}A^T \\ \Rightarrow\ & A^{\dagger}A = \underline{(A^TA)^{-1}}\ \underline{A^TA} = I_n \\ \Rightarrow\ & A^{\dagger}A = I_n \end{aligned} \tag{2.54}$$

(3)

En forma análoga, si se verifica que:

$$rg(A) = m \overset{(\text{th.}2.12)}{\Rightarrow} rg(AA^T) = rg(A^TA) = m$$

Luego, como:

$$\begin{aligned} & A \in \mathcal{M}_{m\times n}(\mathbb{R}) \Rightarrow A^T \in \mathcal{M}_{n\times m}(\mathbb{R}) \Rightarrow AA^T \in \mathcal{M}_{m\times m}(\mathbb{R}) \\ \Rightarrow\ & AA^T \in M_{m\times m}(R) \ \wedge\ rg(AA^T) = m \\ \overset{\text{th.}1.20}{\Rightarrow}\ & \exists(AA^T)^{-1} \\ \overset{(1)}{\Rightarrow}\ & (AA^T)^{\dagger} = (AA^T)^{-1} \\ \overset{(2.54)}{\Rightarrow}\ & A^{\dagger} = A^T(AA^T)^{-1} \\ \Rightarrow\ & AA^{\dagger} = \underline{AA^T}\ \underline{(AA^T)^{-1}} = I_m \\ \Rightarrow\ & AA^{\dagger} = I_m \end{aligned}$$

■

Teorema 2.25.

Sea $A_{m\times n}(\mathbb{R})$. Si $P_{m\times m}(\mathbb{R})$, $Q_{n\times n}(\mathbb{R})$ son matrices ortogonales entonces:

$$(PAQ)^{\dagger} = Q^TA^{\dagger}P^T$$

Demostración:

Por definición se tiene que:

$$X = R^{\dagger} :\Leftrightarrow RXR = R\ ,\ XRX = X\ ,\ (XR)^T = XR\ ,\ (RX)^T = RX \tag{2.55}$$

Por otro lado, como $P_{m\times m}(\mathbb{R})$ y $Q_{n\times n}(\mathbb{R})$ son matrices ortogonales entonces por definición se tiene que:

$$\begin{cases} PP^T = I_m = P^TP \\ QQ^T = I_n = Q^TQ \end{cases} \tag{2.56}$$

Ahora bien tomando $X := Q^TA^\dagger P^T$ y $R := PAQ$, se tiene que:

$$\gtrdot \quad RXR := (PA\underline{Q)(Q^T} A^\dagger \underline{P^T)(P} AQ) \overset{(2.56)}{=} (P\underline{A)A^\dagger(A} Q) := PAQ := R$$

$$\gtrdot \quad XRX := (Q^TA^\dagger \underline{P^T)(P} A\underline{Q)(Q^T} A^\dagger P^T) \overset{(2.56)}{=} (Q^T\underline{A^\dagger)A(A^\dagger} P^T) := Q^TA^\dagger P^T$$

$$\Rightarrow \quad XRX = X$$

$$\gtrdot \quad (XR)^T = [(Q^TA^\dagger \underline{P^T)(P} AQ)]^T \overset{(2.56)}{=} [(Q^T\underline{A^\dagger)(A} Q)]^T := [Q^T(A^\dagger A)^TQ]^T$$

$$\overset{th.}{\Rightarrow} (XR)^T = Q^T(A^\dagger A)Q = Q^T(A^\dagger I_m A)Q \overset{(2.56)}{=} \underline{Q^TA^\dagger(P^T} \underline{P)AQ} := XR$$

$$\gtrdot \quad (RX)^T := [(PA\underline{Q)(Q^T} A^\dagger P^T)]^T \overset{(2.56)}{=} [(P\underline{A)(A^\dagger} P^T)]^T := [P(AA^\dagger)^TP^T]^T$$

$$\overset{th.}{\Rightarrow} (RX)^T = P(AA^\dagger)P^T = P(AI_nA^\dagger)P^T \overset{(2.56)}{=} \underline{PA(Q} \underline{Q^T)A^\dagger P^T} := RX$$

Se concluye por tanto de (2.55) que $R^\dagger = X$, es decir $(PAQ)^\dagger = Q^TA^\dagger P^T$. ■

Observación 2.5. *El resultado que acaba de probarse, indirectamente está diciendo como puede calcularse la inversa de Moore-Penrose usando la descomposición en valores singulares de una matriz A, pues, si:*

$$A = U\Sigma V^T \quad \Rightarrow \quad A^\dagger = (U\Sigma V^T)^\dagger = (V^T)^T\Sigma^\dagger U^T \overset{th.}{=} V\Sigma^\dagger U^T$$

Teorema 2.26.
Dada una matriz $A_{m\times n}(\mathbb{R})$ se cumple que:

1. $(AA^\dagger)^\dagger = AA^\dagger$.
2. $(A^\dagger A)^\dagger = A^\dagger A$.
3. $rg(A) = rg(A^\dagger) = rg(AA^\dagger) = rg(A^\dagger A)$.
4. $AA^\dagger$, $A^\dagger A$, $I_m - AA^\dagger$, $I_n - A^\dagger A$ *son matrices simétricas e idempotentes.*
5. *Si* $rg(A) = r$, *entonces* $\begin{cases} rg(I_m - AA^\dagger) = m - r \\ rg(I_n - A^\dagger A) = n - r \end{cases}$

Demostración:

Bajo las condiciones de la definición se tiene que:

$$X = R^\dagger :\Leftrightarrow RXR = R\ ,\ XRX = X\ ,\ (XR)^T = XR\ ,\ (RX)^T = RX \tag{2.57}$$

(1)

Al considerar $X := AA^\dagger$ y $R := AA^\dagger$, se observa que:

- $RXR := AA^\dagger(A\,\underline{A^\dagger)AA^\dagger} := A\,\underline{A^\dagger(AA^\dagger)} := AA^\dagger := R$
- $XRX := AA^\dagger(A\,\underline{A^\dagger)AA^\dagger} := A\,\underline{A^\dagger(AA^\dagger)} := AA^\dagger := X$
- $(XR)^T := [(A\,\underline{A^\dagger)AA^\dagger}]^T := (AA^\dagger)^T := A\,\underline{A^\dagger} := A(A^\dagger AA^\dagger)$

$\overset{th.}{\Rightarrow} (XR)^T = (AA^\dagger)(AA^\dagger) := XR$

- $(RX)^T := [(A\,\underline{A^\dagger)AA^\dagger}]^T := (AA^\dagger)^T := A\,\underline{A^\dagger} := A(A^\dagger AA^\dagger)$

$\overset{th.}{\Rightarrow} (RX)^T = (AA^\dagger)(AA^\dagger) := RX$

Por tanto se verifican las condiciones de (2.57), esto indica que $R^\dagger = X$, es decir $(AA^\dagger)^\dagger = AA^\dagger$.

(2)

Se puede observar que en el punto anterior, se ha demostrado que $(BB^\dagger)^\dagger = BB^\dagger$, por ende, si:

$$B = A^\dagger \quad \Rightarrow \quad [A^\dagger(A^\dagger)^\dagger]^\dagger = A^\dagger(A^\dagger)^\dagger \quad \overset{\text{th.}2.23}{\Rightarrow} \quad (A^\dagger A)^\dagger = A^\dagger A$$

(3)

Dado que $A := AA^\dagger A$, al usar el teorema 1.14, se tiene que:

$$\begin{cases} rg(A) := rg(A\ \underline{A^\dagger A}) \leq rg(AA^\dagger) \leq rg(A) \\ rg(A) := rg(\ \underline{AA^\dagger}\ A) \leq rg(AA^\dagger) \leq rg(A) \end{cases}$$

Concluyendo por tanto que:

$$rg(A) = rg(A^\dagger A) = rg(AA^\dagger) \tag{2.58}$$

Asimismo, dado que $A^\dagger := A^\dagger A A^\dagger$, al usar nuevamente el teorema 1.14, se obtiene que:

$$\begin{cases} rg(A^\dagger) = rg(A^\dagger \underline{AA^\dagger}) \leq rg(AA^\dagger) \leq rg(A) \\ rg(A^\dagger) = rg(\underline{A^\dagger A}\, A^\dagger) \leq rg(A^\dagger A) \leq rg(A) \end{cases}$$

Infiriendo por tanto que:

$$rg(A^\dagger) = rg(A^\dagger A) = rg(AA^\dagger) \overset{(2.58)}{\Rightarrow} rg(A) = rg(A^\dagger) = rg(A^\dagger A) = rg(AA^\dagger)$$

(4)

$A^\dagger A$ y $AA^\dagger$ son matrices simétricas por definición de la inversa de Moore-Penrose. Ahora, si:

- $B = I_m - AA^\dagger \Rightarrow B^T = (I_m - AA^\dagger)^T \overset{th.}{=} I_m^T - (AA^\dagger)^T = I_m - AA^\dagger = B$
- $C = I_n - A^\dagger A \Rightarrow C^T = (I_n - A^\dagger A)^T \overset{th.}{=} I_n^T - (A^\dagger A)^T = I_m - AA^\dagger = C$

Por lo tanto B y C son matrices simétricas, es decir $I_m - AA^\dagger$ y $I_n - A^\dagger A$ son matrices simétricas. Resta la idempotencia, para esto notar que:

- $(AA^\dagger)^2 := (A\,\underline{A^\dagger)(AA^\dagger)} := AA^\dagger$
- $(A^\dagger A)^2 := (A^\dagger\,\underline{A)(A^\dagger A)} := A^\dagger A$
- $(I_m - AA^\dagger)^2 := (I_m - AA^\dagger)(I_m - AA^\dagger) \overset{th.}{=} (I_m - AA^\dagger)I_m - (I_m - AA^\dagger)AA^\dagger$

$\Rightarrow (I_m - AA^\dagger)^2 = (I_m - AA^\dagger) - [AA^\dagger - (AA^\dagger)^2] = (I_m - AA^\dagger) - (AA^\dagger - AA^\dagger)$

$\Rightarrow (I_m - AA^\dagger)^2 = I_m - AA^\dagger$

- $(I_n - A^\dagger A)^2 = (I_n - A^\dagger A)(I_n - A^\dagger A) = (I_n - A^\dagger A)I_n - (I_n - A^\dagger A)A^\dagger A$

$\Rightarrow (I_n - A^\dagger A)^2 = (I_n - A^\dagger A)I_n - [A^\dagger A - (A^\dagger A)^2] = (I_n - A^\dagger A) - (A^\dagger A - A^\dagger A)$

$\Rightarrow (I_n - A^\dagger A)^2 = I_n - A^\dagger A$

(5)

Dado que $AA^\dagger$, $A^\dagger A$, $I_m - AA^\dagger$, $I_n - A^\dagger A$ son matrices simétricas e idempotentes el teorema 1.41 garantiza que en todas ellas se verifica que el rango coincide con su traza, es decir:

- $rg(I_m - AA^\dagger) = tr(I_m - AA^\dagger) \overset{th.}{=} tr(I_m) - tr(AA^\dagger) = m - rg(AA^\dagger)$

$$\overset{(3)}{\Rightarrow} rg(I_m - AA^\dagger) = m - rg(A) \tag{2.59}$$

Análogamente:

$$\begin{aligned}
&\gtrdot \quad rg(I_n - A^\dagger A) = tr(I_n - A^\dagger A) \overset{th.}{=} tr(I_n) - tr(A^\dagger A) = n - rg(A^\dagger A) \\
&\Rightarrow rg(I_n - A^\dagger A) = n - rg(A)
\end{aligned} \tag{2.60}$$

Ahora, si $rg(A) = r$ entonces de (2.59) y (2.60) se obtiene que $rg(I_m - AA^\dagger) = m - r$ y $rg(I_n - A^\dagger A) = n - r$.

■

Teorema 2.27.

Dada una matriz $A_{m\times n}(\mathbb{R})$ se cumple que:

1. *Si A es una matriz simétrica, entonces $A^\dagger$ es simétrica.*
2. *Si A es simétrica, entonces $AA^\dagger = A^\dagger A$.*

Demostración:

(1)

Para esto, basta notar que, si:

$$B = A^\dagger \;\Rightarrow\; B^T = (A^\dagger)^T \overset{\text{th.}2.23}{=} (A^T)^\dagger := A^\dagger := B$$

Por lo tanto B es una matriz simétrica, es decir $A^\dagger$ es una matriz simétrica.

(2)

Por la definición de inversa de Moore-Penrose se tiene que:

$$AA^\dagger := (AA^\dagger)^T \overset{th.}{=} (A^\dagger)^T A^T \overset{\text{th.}2.23}{=} (A^T)^\dagger A^T := A^\dagger A \;\Rightarrow\; \underline{AA^\dagger = A^\dagger A}$$

■

Teorema 2.28. *Criterios de obtención de la inversa de Moore-Penrose*

Dada una matriz $A_{m\times n}(\mathbb{R})$ se cumple que:

1. *Si $m = n$ y A es una matriz simétrica e idempotente entonces $A^\dagger = A$.*
2. *Si $rg(A) = n$ entonces $A^\dagger = (A^T A)^{-1} A^T$.*
3. *Si $rg(A) = m$ entonces $A^\dagger = A^T (AA^T)^{-1}$.*
4. *Si las columnas de A son ortogonales $AA^T = I_{n\times n}$, entonces $A^\dagger = A^T$.*

Demostración:

(1)

Como A es una matriz simétrica, por el teorema anterior se sabe que:

$$AA^\dagger = A^\dagger A \tag{2.61}$$

Ahora bien, por definición se sabe que:

$$A = \underline{AA^\dagger}\, A \overset{2.61}{=} A^\dagger AA := A^\dagger A^2 := A^\dagger A \tag{2.62}$$

Por otro lado:

$$A^\dagger := A^\dagger\, \underline{AA^\dagger} \overset{2.61}{=} A\, \underline{A^\dagger A} \overset{2.62}{=} AA := A^2 = A \;\Rightarrow\; \underline{A^\dagger = A}|$$

(2)

Como $A_{m\times n}(\mathbb{R})$ por definición se tiene que $A^T \in \mathcal{M}_{n\times m}(\mathbb{R})$, luego por definición de multiplicación se tiene que:

$$AA^T \in \mathcal{M}_{m\times m}(\mathbb{R}) \;\wedge\; A^TA \in \mathcal{M}_{n\times n}(\mathbb{R})$$

Ahora, si:

$$rg(A) = n \overset{\text{th.2.12}}{\Rightarrow} rg(A^TA) = n \;\wedge\; A^TA \in \mathcal{M}_{n\times n}(\mathbb{R}) \overset{\text{th.1.20}}{\Rightarrow} \exists (A^TA)^{-1}$$
$$\overset{\text{th.2.24}}{\Rightarrow} (A^TA)^\dagger = (A^TA)^{-1}$$

Por otro lado, como por el teorema 2.23 se sabe que:

$$A^\dagger = (A^TA)^\dagger A^T = A^T(AA^T)^\dagger \tag{2.63}$$

entonces se deduce que $A^\dagger = (A^TA)^{-1}A^T$.

(3)

En forma similar, si:

$$rg(A) = m \overset{\text{th.2.12}}{\Rightarrow} rg(AA^T) = m \;\wedge\; AA^T \in \mathcal{M}_{m\times m}(\mathbb{R}) \overset{\text{th.1.20}}{\Rightarrow} \exists (AA^T)^{-1}$$
$$\overset{\text{th.2.24}}{\Rightarrow} (AA^T)^\dagger = (AA^T)^{-1}$$

De esto y (2.63) se concluye que $A^\dagger = A^T(AA^T)^{-1}$.

(4)

Como las columnas de A son ortogonales entonces se verifica que $A^TA = I_n$, luego como por (2.63) se tiene que:

$$A^\dagger = (A^TA)^\dagger A^T = (I_n)^\dagger A^T \overset{\text{th.2.24}}{=} I_n^{-1}A^T = A^T \;\Rightarrow\; \boxed{A^\dagger = A^T}$$

■

Teorema 2.29.

Sea $A_{m\times r}(\mathbb{R})$ y sea $B_{r\times n}(\mathbb{R})$. Si $rg(A) = rg(B) = r$ entonces se cumple que:

$$(AB)^\dagger = B^\dagger A^\dagger$$

Demostración:

Como $A \in \mathcal{M}_{m\times r}(\mathbb{R})$ y $rg(A) = r$, el teorema 2.28 garantoza que $A^\dagger = (A^TA)^{-1}A^T$, asimismo, como $B \in \mathcal{M}_{r\times n}(\mathbb{R})$ y $rg(B) = r$ entonces el teorema 2.28 también garantiza que $B^\dagger = B^T(BB^T)^{-1}$.

Considerando lo antes mencionado, se puede definir las siguiente matrices:

$$\begin{cases} A^g := B^\dagger A^\dagger := B^T(BB^T)^{-1}(A^TA)^{-1}A^T \\ A' := AB \end{cases}$$

Afirmación 1: *$A'A^g$ es una matriz simétrica.*

En efecto:

$$A'A^g := (AB)(B^\dagger A^\dagger) := (A\,\underline{B)[B^T(BB^T)^{-1}}\,(A^TA)^{-1}A^T] = A(A^TA)^{-1}A^T$$

$$\Rightarrow (A'A^g)^T = [A(A^TA)^{-1}A^T]^T \overset{th.}{=} (A^T)^T[(A^TA)^{-1}]^TA^T \overset{th.}{=} A[(A^TA)^T]^{-1}A^T$$

$$\overset{\text{th.2.6}}{\Rightarrow} (A'A^g)^T = A(A^TA)^{-1}A^T = A'A^g$$

$$\Rightarrow \boxed{(A'A^g)^T = A'A^g}$$

Afirmación 2: *A^gA' es una matriz simétrica.*

En efecto

$$A^gA' := (B^\dagger A^\dagger)(AB) := [B^T(BB^T)^{-1}\,\underline{(A^TA)^{-1}A^T](A}\,B) = B^T(BB^T)^{-1}B$$

$$\Rightarrow (A^gA')^T = [B^T(BB^T)^{-1}B]^T \overset{th.}{=} B^T[(BB^T)^{-1}]^T(B^T)^T \overset{th.}{=} B^T[(BB^T)^T]^{-1}B$$

$$\overset{\text{th.2.6}}{\Rightarrow} (A^gA')^T = B^T(BB)^{-1}B = A^gA'$$

$$\Rightarrow \boxed{(A^gA')^T = A^gA'}$$

Ahora bien, por lo visto en la afirmación 1 se sabe que:

$$
\begin{aligned}
& A'A^g = A(A^TA)^{-1}A^T \\
\Rightarrow\ & A'A^gA' = [A(A^TA)^{-1}A^T]A' := [A\,\underline{(A^TA)^{-1}A^T]A}\,B = AB := A' \\
\Rightarrow\ & \boxed{A'A^gA' = A'}
\end{aligned}
$$

En forma análoga, por lo visto en la afirmación 2, se tiene que:

$$
\begin{aligned}
& A^gA' = B^T(BB^T)^{-1}B \\
\Rightarrow\ & A^gA'A^g = [B^T(BB^T)^{-1}B]A^g := [B^T(BB^T)^{-1}B]B^\dagger A^\dagger \\
:\Rightarrow\ & A^gA'A^g = [B^T(BB^T)^{-1}\,\underline{B][B^T(BB^T)^{-1}}\,(A^TA)^{-1}A^T] \\
:\Rightarrow\ & A^gA'A^g = [B^T(BB^T)^{-1}][(A^TA)^{-1}A^T] := B^\dagger A^\dagger := A^g \\
\Rightarrow\ & \boxed{A^gA'A^g = A^g}
\end{aligned}
$$

Como puede apreciarse, se ha probado que:

$$
\begin{aligned}
& A'A^gA' = A'\ ,\ A^gA'A^g = A^g\ ,\ (A^gA')^T = A^gA'\ ,\ (A'A^g)^T = A'A^g \\
:\Rightarrow\ & (A')^\dagger = A^g \\
:\Leftrightarrow\ & \boxed{(AB)^\dagger = B^\dagger A^\dagger = B^T(BB^T)^{-1}(A^TA)^{-1}A}
\end{aligned}
$$

■

Teorema 2.30.
Sea $A_{m\times n}(\mathbb{R})$*. Si* $AA^T = A^TA$ *entonces* $AA^\dagger = A^\dagger A$*.*

Demostración:

Como por el teorema 2.23 se sabe que:

$$
\begin{aligned}
& A^\dagger = (A^TA)^\dagger A^T = A^T(AA^T)^\dagger \\
\Rightarrow & \begin{cases} A^\dagger A = (A^TA)^\dagger A^TA = A^T(AA^T)^\dagger A \\ AA^\dagger = A(A^TA)^\dagger A^T = AA^T(AA^T)^\dagger \end{cases}
\end{aligned}
\tag{2.64}
$$

Ahora bien:

$$
\begin{aligned}
& A^\dagger A := (A^\dagger A)^T \overset{(2.64)}{=} [(A^TA)^\dagger A^TA]^T \overset{th.}{=} (A^TA)^T[(A^TA)^\dagger]^T \\
\overset{\text{th.}2.23}{\Rightarrow}\ & A^\dagger A = (A^TA)^T[(A^TA)^T]^\dagger \overset{\text{th.}2.6}{=} (AA^T)(AA^T)^\dagger \overset{(2.64)}{=} AA^\dagger \\
\Rightarrow\ & \boxed{A^\dagger A = AA^\dagger}
\end{aligned}
$$

■

Observación 2.6. *El teorema 2.27 puede verse como un corolario del resultado que acaba de probarse.*

Teorema 2.31.
Dada una matriz $A_{m\times n}(\mathbb{R})$ se cumple que:

1. $R(A) = R(AA^\dagger) = R(AA^T)$.
2. $R(A^\dagger) = R(A^T) = R(A^\dagger A) = R(A^T A)$.
3. $R(I - A^\dagger A) = R(A^T)^\perp = N(A) = N(A^\dagger A)$
4. $R(I - AA^\dagger) = R(A)^\perp = N(A^T) = N(A^\dagger) = N(AA^\dagger)$
5. $R(AB) = R(A) \Leftrightarrow rg(AB) = rg(A), \forall B \in \mathcal{M}_{n\times m}(\mathbb{R})$
6. $N(BA) = N(A) \Leftrightarrow rg(BA) = rg(A), \forall B \in \mathcal{M}_{n\times m}(\mathbb{R})$.

Siendo:

$$\begin{cases} A^\dagger : \text{La inversa de Moore-Penrose de } A \\ R(A) = \{A\overline{x}/\overline{x} \in \mathbb{R}^n\} \\ N(A) = \left\{\overline{x} \in \mathbb{R}^n / A\overline{x} = \overline{0}\right\} \end{cases}$$

Demostración:

(1)

Afirmación 1.1: $R(A) \subset R(AA^\dagger)$

En efecto:

$$\begin{aligned} \forall y \in \mathbb{R}^m : y \in R(A) &:\Rightarrow \exists \overline{x} \in \mathbb{R}^n / y = A\overline{x} \\ &:\Leftrightarrow \ \exists \overline{x} \in \mathbb{R}^n / y = (AA^\dagger A)\overline{x} \\ &\Rightarrow \ \exists \overline{x} \in \mathbb{R}^n / y = AA^\dagger(A\overline{x}) \\ &\Rightarrow \ \exists \overline{z} := A\overline{x} \in \mathbb{R}^m / y = AA^\dagger(\overline{z}) \\ &:\Leftrightarrow \ y \in R(AA^\dagger) \end{aligned}$$

$$\therefore \ \underline{R(A) \subset R(AA^\dagger)}|$$

Afirmación 1.2: $R(AA^\dagger) \subset R(AA^T)$

En efecto:

$$\forall y \in \mathbb{R}^m : y \in R(AA^\dagger) \ :\Rightarrow \ \exists \overline{x} \in \mathbb{R}^m / y = AA^\dagger \overline{x} \tag{2.65}$$

Por otro lado, como por el teorema 2.23 se sabe que:

$$
\begin{aligned}
& A^{\dagger} = (A^T A)^{\dagger} A^T = A^T (AA^T)^{\dagger} && (2.66)\\
& \overset{(2.65)}{\Rightarrow} \exists \overline{x} \in \mathbb{R}^m / y = A[A^T (AA^T)^{\dagger}]\overline{x} \\
& \Rightarrow \exists \overline{x} \in \mathbb{R}^m / y = AA^T[(AA^T)^{\dagger}\overline{x}] \\
& \Rightarrow \exists \overline{z} := (AA^T)^{\dagger}\overline{x} \in \mathbb{R}^m / y = AA^T(\overline{z}) \\
& \Rightarrow y \in R(AA^T) \\
& \therefore \underline{R(AA^{\dagger}) \subset R(AA^T)}|
\end{aligned}
$$

Afirmación 1.3: $R(AA^T) \subset R(A)$

En efecto:

$$
\begin{aligned}
\forall y \in \mathbb{R}^m : y \in R(AA^T) & \Rightarrow \exists \overline{x} \in \mathbb{R}^m / y = AA^T\overline{x} \\
& \Rightarrow \exists \overline{x} \in \mathbb{R}^m / y = A(A^T\overline{x}) \\
& \Rightarrow \exists \overline{z} := A^T\overline{x} \in \mathbb{R}^n / y = A\overline{z} \\
& :\Rightarrow y \in R(A) \\
\therefore \underline{R(AA^T) \subset R(A)}|
\end{aligned}
$$

Como se puede observar de estas tres afirmaciones se concluye que:

$$
\begin{aligned}
& R(A) \subset R(AA^{\dagger}) \subset R(AA^T) \subset R(A) \\
\Rightarrow & \underline{R(A) = R(AA^{\dagger}) = R(AA^T)}|
\end{aligned}
$$

(2)

Afirmación 2.1: $R(A^{\dagger}) \subset R(A^T)$

En efecto:

$$
\begin{aligned}
\forall y \in \mathbb{R}^n : y \in R(A^{\dagger}) & :\Rightarrow \exists \overline{x} \in \mathbb{R}^m / y = A^{\dagger}\overline{x} \\
& \overset{(2.66)}{\Rightarrow} \exists \overline{x} \in \mathbb{R}^m / y = A^T(AA^T)^{\dagger}\overline{x} \\
& \Rightarrow \exists \overline{z} := (AA^T)^{\dagger}\overline{x} \in \mathbb{R}^n / y = A^T\overline{z} \\
& :\Rightarrow y \in R(A^T) \\
\therefore \underline{R(A^{\dagger}) \subset R(A^T)}|
\end{aligned}
$$

Afirmación 2.2: $R(A^T) \subset R(A^{\dagger}A)$

En efecto:

$$
\forall y \in \mathbb{R}^n : y \in R(A^T) :\Rightarrow \exists \overline{x} \in \mathbb{R}^m / y = A^T\overline{x} \tag{2.67}
$$

Ahora , como por el teorema 2.23 se sabe que:

$$
\begin{aligned}
& A^T = A^TAA^\dagger = A^\dagger AA^T \qquad (2.68)\\
\overset{(2.67)}{\Rightarrow} & \exists \overline{x} \in \mathbb{R}^m / y = A^\dagger AA^T\overline{x}\\
\Rightarrow & \exists \overline{z} := A^T\overline{x} \in \mathbb{R}^n / y = A^\dagger A\overline{z}\\
:\Rightarrow & \; y \in R(A^\dagger A)\\
\therefore & \; \underline{R(A^T) \subset R(A^\dagger A)|}
\end{aligned}
$$

Afirmación 2.3: $R(A^\dagger A) \subset R(A^TA)$

En efecto:

$$
\begin{aligned}
& \forall y \in \mathbb{R}^n : y \in R(A^\dagger A)\\
:\Rightarrow & \; \exists \overline{x} \in \mathbb{R}^n / y = A^\dagger A\overline{x}\\
\overset{(2.66)}{\Rightarrow} & \; \exists \overline{x} \in \mathbb{R}^n / y = [A^T(AA^T)^\dagger]A\overline{x}\\
\overset{(2.68)}{\Rightarrow} & \; \exists \overline{x} \in \mathbb{R}^n / y = [(A^TAA^\dagger)(AA^T)^\dagger]A\overline{x}\\
\Rightarrow & \; \exists \overline{z} := A^\dagger(AA^T)^\dagger A \in \mathbb{R}^n / y = A^TA\overline{z}\\
:\Rightarrow & \; y \in R(A^TA)\\
\therefore & \; \underline{R(A^\dagger A) \subset R(A^TA)|}
\end{aligned}
$$

Afirmación 2.4: $R(A^TA) \subset R(A^\dagger)$

En efecto:

$$
\begin{aligned}
\forall y \in \mathbb{R}^n : y \in R(A^TA) \; & :\Rightarrow \; \exists \overline{x} \in \mathbb{R}^n / y = A^TA\overline{x}\\
& \overset{(2.68)}{\Rightarrow} \; \exists \overline{x} \in \mathbb{R}^n / y = (A^\dagger AA^T)A\overline{x}\\
& \Rightarrow \; \exists \overline{z} := AA^TA\overline{x} \in \mathbb{R}^n / y = A^\dagger\overline{z}\\
& :\Rightarrow \; y \in R(A^\dagger)
\end{aligned}
$$

$$\therefore \; \underline{R(A^TA) \subset R(A^\dagger)|}$$

De estas cuatro afirmaciones se tiene que:

$$
\begin{aligned}
& R(A^\dagger) \subset R(A^T) \subset R(A^\dagger A) \subset R(A^TA) \subset R(A^\dagger)\\
\Rightarrow & \; \underline{R(A^\dagger) = R(A^T) = R(A^\dagger A) = R(A^TA)|}
\end{aligned}
$$

(3)

Afirmación 3.1: $R(I - A^{\dagger}A) \subset R(A^T)^{\perp}$

Debe probarse que:

$$\begin{aligned}
\forall y \in \mathbb{R}^n : y \in R(I - A^{\dagger}A) \;&\Rightarrow\; y \in R(A^T)^{\perp} \\
&:\Leftrightarrow\; y \perp R(A^T) \\
&:\Leftrightarrow\; \forall \overline{z} \in R(A^T) : y \perp \overline{z}
\end{aligned}$$

Ahora bien:

$$\begin{aligned}
\forall y \in \mathbb{R}^n : y \in R(I - A^{\dagger}A) \;&\Rightarrow\; \exists \overline{x}_0 \in \mathbb{R}^n / y = (I - A^{\dagger}A)\overline{x}_0 \\
&\Rightarrow\; \exists \overline{x}_0 \in \mathbb{R}^n / y = \overline{x}_0 - A^{\dagger}A\overline{x}_0
\end{aligned}$$

Por otro lado:

$$\forall \overline{z} \in \mathbb{R}^n : \overline{z} \in R(A^T)^{\perp} \;\Rightarrow\; \exists \overline{\xi}_0 \in \mathbb{R}^m / \overline{z} = A^T\overline{\xi}_0$$

Por lo tanto:

$$\begin{aligned}
&\langle \overline{z}, y\rangle = \left\langle A^T\overline{\xi}_0, \overline{x}_0 - A^{\dagger}A\overline{x}_0 \right\rangle = (A^T\overline{\xi}_0)^T \cdot (\overline{x}_0 - A^{\dagger}A\overline{x}_0) \overset{th.}{=} (\overline{\xi}_0{}^T A)(\overline{x}_0 - A^{\dagger}A\overline{x}_0) \\
&\Rightarrow \langle \overline{z}, y\rangle = \overline{\xi}_0{}^T \cdot (A\overline{x}_0 - AA^{\dagger}A\overline{x}_0) := \overline{\xi}_0{}^T \cdot (A\overline{x}_0 - A\overline{x}_0) = \left\langle \overline{\xi}_0, A\overline{x}_0 - A\overline{x}_0 \right\rangle \\
&\Rightarrow \langle \overline{z}, y\rangle = \left\langle \overline{\xi}_0, \overline{0} \right\rangle = 0 \;\;;\;\; \forall \overline{z} \in R(A^T) \\
&:\Rightarrow y \perp \overline{z}, \forall \overline{z} \in R(A^T) \\
&\Rightarrow y \perp R(A^T) \\
&:\Rightarrow y \in R(A^T)^{\perp}
\end{aligned}$$

De este modo se ha probado que:

$$\begin{aligned}
&\forall y \in \mathbb{R}^n : y \in R(I - A^{\dagger}A) \;\Rightarrow\; y \in R(A^T)^{\perp} \\
\Rightarrow\;& \underline{R(I - A^{\dagger}A) \subset R(A^T)^{\perp}}\big|
\end{aligned}$$

Afirmación 3.2: $R(A^T)^{\perp} \subset N(A)$

Debe probarse que dado $y \in R(A^T)^{\perp}$, se verifica que $y \in N(A)$. Para esto notar que:

$$\forall y \in \mathbb{R}^n : y \in R(A^T)^{\perp} \;\Rightarrow\; y \perp R(A^T) \;\Leftrightarrow\; \forall \overline{\xi} \in R(A^T) : y \perp \overline{\xi} \tag{2.69}$$

Por otro lado, como:

$$y \in \mathbb{R}^n \;\Rightarrow\; \exists \overline{z}_0 := Ay \in \mathbb{R}^m$$

Ahora bien:

$$\begin{aligned} &\forall \overline{w} \in \mathbb{R}^m : \langle \overline{w}, \overline{z}_0 \rangle = \langle \overline{w}, Ay \rangle = \overline{w}^T \cdot Ay \overset{th.}{=} \overline{w}^T \cdot [(A^T)^T y] \overset{th.}{=} (A^T \overline{w})^T \cdot y \\ \Rightarrow &\forall \overline{w} \in \mathbb{R}^m : \langle \overline{w}, \overline{z}_0 \rangle = \left\langle A^T \overline{w}, y \right\rangle \end{aligned} \tag{2.70}$$

Luego como se sabe que:

$$\begin{aligned} & R(A^T) := \{A^T \overline{b} \,/\, \overline{b} \in \mathbb{R}^m\} \;\Rightarrow\; A^T \overline{w} \in R(A^T) \overset{(2.69)}{\Rightarrow} y \perp (A^T \overline{w}) \\ :\Leftrightarrow\; & \left\langle y, A^T \overline{w} \right\rangle = 0 \\ \overset{(2.70)}{\Rightarrow}\; & \langle \overline{w}, \overline{z}_0 \rangle = 0, \forall \overline{w} \in \mathbb{R}^m \\ \Rightarrow\; & \langle \overline{w}, Ay \rangle = 0, \forall \overline{w} \in \mathbb{R}^m \\ \Rightarrow\; & \langle Ay, Ay \rangle = 0 \\ :\Rightarrow\; & Ay = \overline{0} \\ :\Rightarrow\; & y \in N(A) \end{aligned}$$

De este modo queda probado que:

$$\begin{aligned} &\forall y \in \mathbb{R}^n : y \in R(A^T)^\perp \;\Rightarrow\; y \in N(A) \\ \Rightarrow\; & \underline{R(A^T)^\perp \subset N(A)|} \end{aligned}$$

Afirmación 3.3: $N(A) \subset N(A^\dagger A)$

Para esto observar que:

$$\begin{aligned} \forall \overline{x} \in \mathbb{R}^n : \overline{x} \in N(A) \;&\Rightarrow A\overline{x} = \overline{0} \\ &\Rightarrow A^\dagger A \overline{x} = A^\dagger \overline{0} = \overline{0} \\ &\Rightarrow A^\dagger A \overline{x} = \overline{0} \\ &:\Leftrightarrow \overline{x} \in N(A^\dagger A) \\ \Rightarrow\; \underline{N(A) \subset N(A^\dagger A)|} & \end{aligned}$$

Afirmación 3.4: $N(A^\dagger A) \subset R(I - A^\dagger A)$

Por definición se sabe que:

$$R(I - A^\dagger A) = \{(I - A^\dagger A)\overline{b} \,/\, \overline{b} \in \mathbb{R}^m\} \tag{2.71}$$

Ahora:

$$\begin{aligned}
\forall \overline{x} \in \mathbb{R}^m : \overline{x} \in N(A^\dagger A) &\Rightarrow A^\dagger A\overline{x} = \overline{0} \\
&\Rightarrow \overline{x} - A^\dagger A\overline{x} = \overline{x} - \overline{0} = \overline{x} \\
&\Rightarrow \overline{x} - A^\dagger A\overline{x} = \overline{x} \\
&\Rightarrow \overline{x} = (I - A^\dagger A)\overline{x} \\
&\Rightarrow \exists \overline{b} := \overline{x} \in \mathbb{R}^m / \overline{x} = (I - A^\dagger A)\overline{b} \\
&\overset{(2.71)}{\Rightarrow} \overline{x} \in R(I - A^\dagger A)
\end{aligned}$$

Por tanto de estas 4 afirmaciones se tiene que:

$$\begin{aligned}
& R(I - A^\dagger A) \subset R(A^T)^\perp \subset N(A) \subset N(A^\dagger A) \subset R(I - A^\dagger A) \\
\Rightarrow\ & \underline{R(I - A^\dagger A) = R(A^T)^\perp = N(A) = N(A^\dagger A)}\Big|
\end{aligned}$$

(4)

Afirmación 4.1: $R(I - AA^\dagger) \subset R(A)^\perp$

Debe probarse que:

$$\begin{aligned}
\forall y \in \mathbb{R}^m : y \in R(I - AA^\dagger) &\Rightarrow y \in R(A)^\perp \\
&:\Leftrightarrow y \perp R(A) \\
&:\Leftrightarrow \forall \xi \in R(A) : y \perp \xi
\end{aligned} \tag{2.72}$$

Para esto notar que, dado $y \in \mathbb{R}^m$, si:

$$\begin{aligned}
y \in R(I - AA^\dagger) &\Rightarrow \exists\, \overline{x}_0 \in \mathbb{R}^m / y = (I - AA^\dagger)\overline{x}_0 \\
&\Rightarrow \exists\, \overline{x}_0 \in \mathbb{R}^m / y = \overline{x}_0 - AA^\dagger\overline{x}_0
\end{aligned}$$

Por otro lado, dado $\overline{\xi} \in \mathbb{R}^m$ se verifica que, si:

$$\overline{\xi} \in R(A) \ :\Rightarrow\ \exists\, \overline{z}_0 \in \mathbb{R}^n / \overline{\xi} = A\overline{z}_0$$

Ahora bien:

$$\begin{aligned}
\left\langle \overline{\xi}, y \right\rangle = \langle A\overline{z}_0, y \rangle &= \left\langle A\overline{z}_0, \overline{x}_0 - AA^\dagger\overline{x}_0 \right\rangle \\
&= (A\overline{z}_0)^T \cdot (\overline{x}_0 - AA^\dagger\overline{x}_0) \\
&= (\overline{z}_0{}^T A^T)(\overline{x}_0 - AA^\dagger\overline{x}_0) \\
&= (\overline{z}_0)^T \cdot (A^T\overline{x}_0 - A^T AA^\dagger\overline{x}_0) \\
&= \left\langle \overline{z}_0, A^T\overline{x}_0 - A^T AA^\dagger\overline{x}_0 \right\rangle
\end{aligned}$$

Lo anterior implica que:

$$\begin{aligned}
\langle \overline{\xi}, y\rangle &\overset{(2.68)}{=} \langle \overline{z}_0, A^T\overline{x}_0 - A^T\overline{x}_0\rangle = \langle \overline{z}_0, \overline{0}\rangle \\
&\qquad\qquad\qquad\qquad\quad = 0 \\
\Rightarrow\ & y \perp \overline{\xi}\,, \forall \overline{\xi} \in R(A) \\
\Rightarrow\ & y \perp R(A)
\end{aligned}$$

Así (2.72) queda probado.

Afirmación 4.2: $R(A)^{\perp} \subset N(A^T)$

Se debe probar que:

$$\forall y \in \mathbb{R}^m : y \in R(A)^{\perp} \ \Rightarrow\ y \in N(A^T) \ \Leftrightarrow\ A^T y = \overline{0} \tag{2.73}$$

Asumiendo que y es un vector cualquiera de $\mathbb{R}^m$ se tiene que, si:

$$\begin{aligned}
y \in R(A)^{\perp} \ :\Rightarrow\ y \perp R(A) \ &:\Leftrightarrow\ \forall \overline{z} \in R(A) : y \perp z \\
&\Leftrightarrow\ \forall \overline{z} \in R(A) : \langle y, z\rangle = 0
\end{aligned}$$

Ahora bien:

$$\begin{aligned}
\forall \overline{\xi} \in \mathbb{R}^n : A\overline{\xi} \in R(A) \ &\Rightarrow\ \langle y, A\overline{\xi}\rangle = 0 \\
&\Rightarrow\ 0 = \langle y, A\overline{\xi}\rangle = y^T \cdot (A\overline{\xi}) \\
&\quad\ \overset{th.}{=}\ y^T \cdot [(A^T)^T\overline{\xi}] \\
&\quad\ \overset{th.}{=}\ (A^T y)^T \cdot \overline{\xi} \\
&\quad\ = \langle A^T y, \overline{\xi}\rangle \\
&\Rightarrow\ \langle A^T y, \overline{\xi}\rangle = 0, \forall \overline{\xi} \in \mathbb{R}^n \\
&\Rightarrow\ \langle A^T y, A^T y\rangle = 0 \\
&\Rightarrow\ A^T y = \overline{0} \\
&:\Rightarrow\ y \in N(A^T)
\end{aligned}$$

De este modo (2.73) queda probado.

Afirmación 4.3: $N(A^T) \subset N(A^{\dagger})$

En efecto:

$$\forall \overline{x} \in \mathbb{R}^m : \overline{x} \in N(A^T) \ :\Rightarrow\ A^T\overline{x} = \overline{0} \tag{2.74}$$

Ahora, como por (2.66) se sabe que:

$$
\begin{aligned}
A^{\dagger} = (A^T A)^{\dagger} A^T \ &\Rightarrow\ A^{\dagger}\overline{x} = (A^T A)^{\dagger}\ \underline{A^T\overline{x}} \\
&\overset{(2.74)}{\Rightarrow}\ A^{\dagger}\overline{x} = (A^T A)^{\dagger}\overline{0} \\
&\Rightarrow\ A^{\dagger}\overline{x} = \overline{0} \\
&:\Rightarrow\ \overline{x} \in N(A^{\dagger})
\end{aligned}
$$

Por lo tanto

$$
\begin{aligned}
\forall \overline{x} \in \mathbb{R}^m : \overline{x} \in N(A^T) \ &\Rightarrow\ \overline{x} \in N(A^{\dagger}) \\
\Rightarrow\ \underline{N(A^T) \subset N(A^{\dagger})}\Big| &
\end{aligned}
$$

Afirmación 4.4: $N(A^{\dagger}) \subset N(AA^{\dagger})$

En efecto:

$$
\begin{aligned}
\forall \overline{x} \in \mathbb{R}^m : \overline{x} \in N(A^{\dagger}) \ &:\Rightarrow\ A^{\dagger}\overline{x} = \overline{0} \\
&\Rightarrow\ AA^{\dagger}\overline{x} = A\overline{0} = \overline{0} \\
&\Rightarrow\ AA^{\dagger}\overline{x} = \overline{0} \\
&\Rightarrow\ \overline{x} \in N(AA^{\dagger}) \\
\Rightarrow\ \underline{N(A^{\dagger}) \subset N(AA^{\dagger})}\Big| &
\end{aligned}
$$

Afirmación 4.5: $N(AA^{\dagger}) \subset R(I - AA^{\dagger})$

En efecto:

$$
\begin{aligned}
\forall \overline{x} \in \mathbb{R}^m : \overline{x} \in N(AA^{\dagger}) \ &:\Rightarrow\ AA^{\dagger}\overline{x} = \overline{0} \\
&\Rightarrow\ \overline{x} - AA^{\dagger}\overline{x} = \overline{x} - \overline{0} = \overline{x} \\
&\Rightarrow\ \overline{x} - AA^{\dagger}\overline{x} = \overline{x} \\
&\Rightarrow\ \overline{x} = (I - AA^{\dagger})\overline{x} \\
&\Rightarrow\ \exists \overline{z} \in \mathbb{R}^m := \overline{x} / \overline{x} = (I - AA^{\dagger})\overline{z} \\
&:\Rightarrow\ \overline{x} \in R(I - AA^{\dagger}) \\
\Rightarrow\ \underline{N(AA^{\dagger}) \subset R(I - AA^{\dagger})}\Big| &
\end{aligned}
$$

De estas 5 afirmaciones se observa que:

$$
\begin{aligned}
& R(I - AA^{\dagger}) \subset R(A)^{\perp} \subset N(A^T) \subset N(A^{\dagger}) \subset N(AA^{\dagger}) \subset R(I - AA^{\dagger}) \\
\Rightarrow\ & \underline{R(I - AA^{\dagger}) = R(A)^{\perp} = N(A^T) = N(A^{\dagger}) = N(AA^{\dagger})}\Big|
\end{aligned}
$$

(5)

$(\Rightarrow)$

Si:

$$\begin{aligned} R(AB) = R(A) \;&:\Rightarrow\; Col(AB) = Col(A) \\ &\Rightarrow\; \dim[Col(AB)] = \dim[Col(A)] \\ &:\Leftrightarrow\; rg(AB) = rg(A) \end{aligned}$$

$(\Leftarrow)$

Por hipótesis se tiene que:

$$rg(AB) = rg(A) \;:\Rightarrow\; \dim[Col(AB)] = \dim[Col(A)] \tag{2.75}$$

Ahora bien, por definición se sabe que:

$$Col(AB)\,\tilde{\subset}\,\mathbb{R}^m \quad \wedge \quad Col(A)\,\tilde{\subset}\,\mathbb{R}^m$$

Afirmación 5.1: $R(AB) \subset R(A)$

En efecto:

$$\begin{aligned} \forall y \in R^m : y \in R(AB) \;&:\Rightarrow\; \exists \overline{x}_0 \in R^m / y = AB\overline{x}_0 \\ &:\Rightarrow\; \exists \overline{z}_0 := B\overline{x}_0 \in R^n / y = A\overline{z}_0 \\ &:\Rightarrow\; y \in R(A) \end{aligned}$$

$$\Rightarrow \;\; \underline{R(AB) \subset R(A)|}$$

Lo anterior está diciendo que:

$$\begin{aligned} Col(AB)\,\tilde{\subset}\,Col(A) \;&\overset{(2.75)}{\Rightarrow}\; Col(AB)\,\tilde{\subset}\,Col(A) \;\wedge\; \dim[Col(AB)] = \dim[Col(A)] \\ &\Rightarrow\; Col(AB) = Col(A) \\ &\Rightarrow\; R(AB) = R(A) \end{aligned}$$

(6)

$(\Rightarrow)$

Si:

$$N(BA) = N(A) \;\Rightarrow\; \dim[N(BA)] = \dim[N(A)] \tag{2.76}$$

Por otro lado, como $A_{m\times n}(\mathbb{R})$ y $(BA)_{n\times n}(\mathbb{R})$, el teorema 1.19 garantiza que:

$$rg(A) + \dim[N(A)] = n \;\wedge\; rg(BA) + \dim[N(BA)] = n$$

Pero dada la igualdad en (2.76) se concluye que $rg(BA) = rg(A)$.

($\Leftarrow$)

Al igual que se hizo en ítem *5*, por definición se sabe que:

$$N(A) \tilde{\subset} \mathbb{R}^n \ \wedge \ N(BA) \tilde{\subset} \mathbb{R}^n$$

Afirmación 6.1: $N(A) \subset N(AB)$

En efecto:

$$\begin{aligned} \forall \overline{x} \in \mathbb{R}^n : \overline{x} \in N(A) \ &:\Rightarrow \ A\overline{x} = \overline{0} \\ &\Rightarrow \ BA\overline{x} = B\overline{0} = \overline{0} \\ &\Rightarrow \ BA\overline{x} = \overline{0} \\ &:\Rightarrow \ \overline{x} \in N(BA) \\ \Rightarrow \ \underline{N(A) \subset N(BA)}\Big| & \end{aligned}$$

Ahora bien, por lo mencionado anteriormente y por lo que se tiene como hipótesis, se tiene que:

$$rg(A) + \dim[N(A)] = n = rg(BA) + \dim[N(BA)]$$

Por lo tanto, es claro que $\dim[N(A)] = \dim[N(BA)]$ y como por la afirmacion 6.1 se sabe que $N(A) \subset N(AB)$, entonces se concluye que $N(A) = N(AB)$.

■

Teorema 2.32.

Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $A^+ \in \mathcal{M}_{n\times m}(\mathbb{R})$. La matriz A^+ es la inversa de Moore-Penrose de A, si y sólo si verifica las siguientes condiciones.

1. *AA^+ es la matriz proyección ortogonal sobre $Col(A) \subset \mathbb{R}^m$ respecto a la base usual de $\mathbb{R}^m$.*
2. *A^+A es la matriz proyección ortogonal sobre $Col(A^+) \subset \mathbb{R}^n$ respecto a la base usual de $\mathbb{R}^n$.*

Demostración:

($\Rightarrow$)

Asumiendo que A^+ es la inversa de Moore-Penrose de la matriz A, debe probarse las condición *1* y *2*.

(1)

Como A^+ es la inversa de Moore-Penrose, el teorema 2.26 garantiza que AA^+ es una matriz simétrica e idempotente, por tanto el teorema 1.56 señala que AA^+ es una matriz de proyección ortogonal sobre su espacio columna.

Por otro lado, por el teorema 2.31 se tiene que:

$$R(AA^+) = R(A) \;:\Leftrightarrow\; Col(AA^+) = Col(A)$$

Por lo tanto AA^+ es una matriz de proyección ortogonal sobre $Col(A) \,\tilde{\subset}\, \mathbb{R}^m$.

(2)

Como A^+ es la inversa de Moore-Penrose, el teorema 2.26 garantiza que A^+A es una matriz simétrica e idempotente, por tanto el teorema 1.56 señala que A^+A es una matriz de proyección ortogonal sobre su espacio columna.

Por otro lado, por el teorema 2.31 se tiene que:

$$R(A^+A) = R(A^+) \;:\Leftrightarrow\; Col(A^+A) = Col(A^+)$$

Por lo tanto A^+A es una matriz de proyección ortogonal sobre $Col(A^+) \,\tilde{\subset}\, \mathbb{R}^n$.

$(\Leftarrow)$

Asumiendo que $A^+ \in \mathcal{M}_{n\times m}(\mathbb{R})$ es una matriz que hace que se verifique *1* y *2*, debe probarse que A^+ es la inversa de Moore-Penrose de la matriz A.

Afirmación 1. AA^+ *es una matriz simétrica.*

En efecto, como AA^+ es una matriz proyección ortogonal, por teorema se sabe que AA^+ es una matriz simétrica.

Afirmación 2. A^+A *es una matriz simétrica.*

En efecto, como A^+A es una matriz proyección ortogonal, por teorema se sabe que A^+A es una matriz simétrica.

Afirmación 3. $AA^+A = A$.

Si se considera que $M := AA^+$, entonces M es una matriz proyección ortogonal sobre $Col(A)$.

Bajo las condiciones anteriores, el teorema 1.55 garantiza que:

$$\forall \overline{z} \in Col(A) : M\overline{z} = \overline{z} \tag{2.77}$$

Ahora bien:

$$\begin{aligned}
&\forall \overline{x} \in \mathbb{R}^n : MA\overline{x} = M(A\overline{x}) \overset{(2.77)}{=} A\overline{x} \\
\Rightarrow\ &\forall \overline{x} \in \mathbb{R}^n : MA\overline{x} = A\overline{x} \\
\Rightarrow\ &\forall \overline{x} \in \mathbb{R}^n : (MA - A)\overline{x} = \overline{0} \\
\Rightarrow\ &MA - A = \Theta \\
\Rightarrow\ &MA = A \\
\Rightarrow\ &\underline{AA^+A = A|}
\end{aligned}$$

Afirmación 4. $A^+AA^+ = A^+$.

Si se considera que $M' := A^+A$, entonces M' es una matriz proyección ortogonal sobre $Col(A^+)$.

Bajo las condiciones anteriores el teorema 1.55 garantiza que:

$$\forall \overline{z} \in Col(A^+) : M'\overline{z} = \overline{z} \tag{2.78}$$

Ahora bien:

$$\begin{aligned}
&\forall \overline{x} \in \mathbb{R}^m : M'A^+\overline{x} = M'(A^+\overline{x}) \overset{(2.78)}{=} A^+\overline{x} \\
\Rightarrow\ &\forall \overline{x} \in \mathbb{R}^m : M'A^+\overline{x} = A^+\overline{x} \\
\Rightarrow\ &\forall \overline{x} \in \mathbb{R}^m : (M'A^+ - A^+)\overline{x} = \overline{0} \\
\Rightarrow\ &M'A^+ - A^+ = \Theta \\
\Rightarrow\ &A^+AA^+ = A^+ \\
\Rightarrow\ &\underline{A^+AA^+ = A^+|}
\end{aligned}$$

Por tanto las 4 afirmaciones antes probadas, garantizan que A^+ es inversa de Moore-Penrose de la matriz A.

■

Nota 2.2. *La definición de inversa de Moore-Penrose presentada en este trabajo, es la definición de inversa generalizada que presentó Penrose en un inicio, y la definición de inversa generalizada que presentó Moore es la de aquella matriz que satisface las condiciones 1 y 2 del teorema que acaba de probarse, sin embargo, dado que ambas definiciones son equivalentes tal y como acaba de probarse, a dicha inversa se le otorga el nombre de inversa de Moore-Penrose.*

Teorema 2.33.

Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y sea $T : \mathbb{R}^n \to \mathbb{R}^m$ una aplicación lineal.

Si A es la matriz estándar de la aplicación lineal T entonces la aplicación:

$$\begin{array}{rcl} \Phi : Ker(T)^{\perp} & \longrightarrow & Im(T) \\ \overline{x} & \mapsto & \Phi(\overline{x}) := T(\overline{x}),\ \forall\, \overline{x} \end{array}$$

es un isomorfismo, es decir $\Phi := T\Big|_{Ker(T)^{\perp}}$ es un isomorfismo.

Demostración:

Afirmación 1. $\mathbb{R}^n = Ker(T)^{\perp} \oplus Ker(T)$.

En efecto, como $T : \mathbb{R}^n \to \mathbb{R}^m$ es una aplicación lineal, entonces por el corolario 1.3 se sabe que $Ker(T)$ es un subespacio de $\mathbb{R}^n$, por lo tanto, en base a lo que se mencionó en la observación 1.2 se tiene que $\mathbb{R}^n = Ker(T)^{\perp} \oplus Ker(T)$ y por tanto de acuerdo a lo mencionado en el teorema 1.2 se tiene que:

$$n = \dim[Ker(T)^{\perp}] + \dim[Ker(T)] \tag{2.79}$$

Afirmación 2. $\mathbb{R}^m = \mathrm{Im}(T) \oplus \mathrm{Im}(T)^{\perp}$.

En efecto, como $T : \mathbb{R}^n \to \mathbb{R}^m$ es una aplicación lineal, entonces por el corolario 1.3 se sabe que $Ker(T)$ es un subespacio de $\mathbb{R}^m$, por lo tanto, en base al teorema 1.3 se tiene que $\mathbb{R}^m = \mathrm{Im}(T) \oplus \mathrm{Im}(T)^{\perp}$ y por ende, de acuerdo a lo mencionado en el teorema 1.2 se tiene que:

$$m = \dim[\mathrm{Im}(T)] + \dim[\mathrm{Im}(T)^{\perp}]$$

Afirmación 3. Φ *es una aplicación lineal.*

En efecto lo es, por ser la restricción de una aplicación lineal a un subespacio de su dominio.

Afirmación 4. Φ *es una aplicación inyectiva.*

En efecto, sean $\overline{x}_1$ y $\overline{x}_2$ un par de vectores de $Ker(T)^{\perp}$, si:

$$\begin{aligned} & \Phi(\overline{x}_1) = \Phi(\overline{x}_2) \ \Rightarrow\ \Phi(\overline{x}_1) - \Phi(\overline{x}_2) = \overline{0} \ \Rightarrow\ \Phi(\overline{x}_1 - \overline{x}_2) = \overline{0} \\ :\Rightarrow\ & T(\overline{x}_1 - \overline{x}_2) = \overline{0} \ :\Rightarrow\ (\overline{x}_1 - \overline{x}_2) \in Ker(T) \\ \Rightarrow\ & (\overline{x}_1 - \overline{x}_2) \in Ker(T) \ \wedge\ (\overline{x}_1 - \overline{x}_2) \in Ker(T)^{\perp} \\ :\Rightarrow\ & (\overline{x}_1 - \overline{x}_2) \in Ker(T) \cap Ker(T)^{\perp} \end{aligned}$$

Ahora bien, por el teorema 1.2 se tiene que $(\overline{x}_1 - \overline{x}_2) \in \{\overline{0}\}$ lo cual equivale a decir que $\overline{x}_1 = \overline{x}_2$, es decir Φ es una aplicación lineal inyectiva.

Afirmación 5. $\dim[Ker(T)^{\perp}] = \dim[\mathrm{Im}(T)]$

Por el teorema 1.53 se sabe que:

$$\begin{cases} Ker(T) = Nul(A) \\ rg(T) = rg(A) \end{cases} \Rightarrow \begin{cases} \dim[Ker(T)] = \dim[Nul(A)] \\ \dim[\mathrm{Im}(T)] = rg(A) \end{cases} \tag{2.80}$$

Ahora, como $A \in \mathbb{M}_{m\times n}(\mathbb{R})$ por el teorema del rango (th. 1.19), se tiene que:

$$\begin{aligned}
& \dim[col(A)] + \dim[Nul(A)] = n \\
\overset{(2.79)}{\Rightarrow}\ & \dim[col(A)] + \dim[Nul(A)] = \dim[Ker(T)^{\perp}] + \dim[Ker(T)] \\
\overset{(2.80)}{\Rightarrow}\ & \dim[col(A)] + \dim[Ker(T)] = \dim[Ker(T)^{\perp}] + \dim[Ker(T)] \\
\Rightarrow\ & \dim[col(A)] = \dim[Ker(T)^{\perp}] \\
:\Rightarrow\ & rg(A) = \dim[Ker(T)^{\perp}] \\
\overset{(2.80)}{\Rightarrow}\ & \dim[\mathrm{Im}(T)] = \dim[Ker(T)^{\perp}]
\end{aligned}$$

Así, de las afirmaciones 3, 4 y 5 se sigue que Φ es un isomorfismo.

■

Teorema 2.34.
Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $T : \mathbb{R}^n \to \mathbb{R}^m$ una aplicación lineal y sea $\Phi := T\Big|_{Ker(T)^{\perp}}$. Si A es la matriz estándar de la aplicación lineal T entonces la matriz estándar de la aplicación lineal

$$\begin{aligned} \tilde{T} : \mathbb{R}^m &\longrightarrow \mathbb{R}^n \\ \overline{x} &\mapsto \tilde{T}(\overline{x}) := \Phi^{-1} \circ Proy_{\mathrm{Im}(T)}(\overline{x}), \ \forall \overline{x} \end{aligned}$$

es la matriz $A^{\dagger}$, es decir, es la inversa de Moore-Penrose de la matriz A.

Demostración:

Dado que $\Phi := T\Big|_{Ker(T)^{\perp}}$, se tiene que:

$$\begin{aligned} \Phi : Ker(T)^{\perp} &\longrightarrow Im(T) \\ \overline{x} &\mapsto \Phi(\overline{x}) := T(\overline{x}), \ \forall \overline{x} \end{aligned}$$

Es un isomorfismo, tal y como se demostró en el teorema anterior, esto quiere decir que existe la aplicación $\Phi^{-1} : Im(T) \longrightarrow Ker(T)^{\perp}$, que naturalmente también es un isomorfismo.

Afirmación 1. *$\tilde{T}$ es un operador lineal.*

En efecto, por el teorema 1.54 se sabe que $Proy_{Im(T)}$ es una aplicación lineal y como por lo antes mencionado Φ^{-1} también es una aplicación, entonces es claro que $\tilde{T}$ también es una aplicación lineal por ser composición de aplicaciones lineales.

Afirmación 2. $Im(\tilde{T}) = Ker(T)^{\perp}$.

En efecto:

$$\begin{aligned} Im(\tilde{T}) &:= \tilde{T}(\mathbb{R}^m) \\ &:= \Phi^{-1} \circ Proy_{Im(T)}(\mathbb{R}^m) \\ &:= \Phi^{-1}\left[Proy_{Im(T)}(\mathbb{R}^m)\right] \\ &:= \Phi^{-1}(Im(T)) \\ &:= Ker(T)^{\perp} \end{aligned}$$

Afirmación 3. Si $v_1 = Proy_{Im(T)}(v) \;\Rightarrow\; T \circ \tilde{T}(v) = v_1$.

En efecto, si se asume que $v_1 = Proy_{Im(T)}(v)$ por definición se tiene que:

$$\begin{aligned} T \circ \tilde{T}(v) &= T[\tilde{T}(v)] \\ &:= T[\Phi^{-1} \circ Proy_{Im(T)}(v)] \\ &= T\left\{\Phi^{-1}\left[Proy_{Im(T)}(v)\right]\right\} \\ &= T\left\{\Phi^{-1}\left[Proy_{Im(T)}\left(Proy_{Im(T)}(v)\right)\right]\right\} \\ &\overset{\text{th.1.54}}{=} T\left\{\Phi^{-1}\left[Proy_{Im(T)}(v)\right]\right\} \\ &= T[\underbrace{\Phi^{-1}(v_1)}_{\in Ker(T)^{\perp}}] \\ &= \Phi[\Phi^{-1}(v_1)] \\ &:= v_1 \end{aligned}$$

$$\Rightarrow T \circ \tilde{T}(v) = Proy_{Im(T)}(v), \forall v$$

$$:\Rightarrow T \circ \tilde{T} = Proy_{Im(T)}$$

Afirmación 4. *Si* $u_1 = Proy_{Ker(T)^{\perp}}(u)$, *entonces* $\tilde{T} \circ T(u) = u_1$.

En efecto, dado $u \in \mathbb{R}^n$ se tiene que:

$$\begin{aligned} \tilde{T} \circ T(u) &:= \tilde{T}[T(u)] \\ &:= \Phi^{-1} \circ Proy_{Im(T)}(T(u)) \\ &\overset{\text{th.1.54}}{=} \Phi^{-1}[T(u)] \end{aligned}$$

Ahora bien, como:

$$u_1 = Proy_{Ker(T)^{\perp}}(u) \;:\Rightarrow\; \begin{cases} u_1 \in Ker(T)^{\perp} \\ u_1 - u \perp Ker(T)^{\perp} \end{cases}$$

Así, dado que:

$$u_1 \in Ker(T)^{\perp} \;:\Rightarrow\; T(u_1) = \Phi(u_1) \;\Rightarrow\; \Phi^{-1}[T(u_1)] = u_1 \tag{2.81}$$

Asimismo, como:

$$\begin{aligned} u_1 - u \perp Ker(T)^{\perp} &:\Rightarrow u_1 - u \in [Ker(T)^{\perp}]^{\perp} \\ &\overset{\text{th.1.2}}{\Rightarrow} u_1 - u \in Ker(T) \\ &:\Rightarrow T(u_1 - u) = \overline{0} \\ &\Rightarrow T(u_1) = T(u) \\ &\overset{(2.81)}{\Rightarrow} \Phi^{-1}[T(u)] = u_1 \\ &\Rightarrow \tilde{T} \circ T(u) = Proy_{Ker(T)^{\perp}}(u), \forall u \\ &:\Rightarrow \tilde{T} \circ T = Proy_{Ker(T)^{\perp}} = Proy_{Im(\tilde{T})} \end{aligned}$$

En conclusión, como:

$$\begin{cases} \tilde{T} \circ T = Proy_{Ker(T)^{\perp}} = Proy_{Im(\tilde{T})} \\ T \circ \tilde{T} = Proy_{Im(T)} \end{cases}$$

$$\overset{\text{th.1.49}}{\Rightarrow} \begin{cases} \left[Proy_{Ker(T)^{\perp}}\right] = \left[Proy_{Im(\tilde{T})}\right] = [\tilde{T} \circ T] = [\tilde{T}][T] \\ \left[Proy_{Im(T)}\right] = [T \circ \tilde{T}] = [T][\tilde{T}] \end{cases}$$

Por definición, se sabe que $\left[Proy_{Im(T)}\right] =: P_1$ es la matriz de proyección ortogonal sobre $Im(T) \overset{\text{th.1.53}}{=} Col(A)$.

Asimismo, si $[\tilde{T}] =: \tilde{A}$, es decir si $\tilde{A}$ es matriz estándar de $\tilde{T}$, por definición la matriz $\left[Proy_{Im(\tilde{T})}\right] =: P_2$ es la proyección ortogonal sobre $Im(\tilde{T}) \overset{\text{th.1.53}}{=} Col(\tilde{A})$.

Por lo tanto se tendría que $\tilde{A}$ es aquella matriz que hace que, $A\tilde{A}$ sea la matriz proyección ortogonal sobre $Col(A)$ y $\tilde{A}A$ sea la matriz proyección ortogonal sobre $Col(\tilde{A})$, bajo estas condiciones el teorema 2.32 garantiza que $\tilde{A} = A^{\dagger}$.

■

Nota 2.3. *Adecuando la definición de la transformación lineal de Moore-Penrose que aparece en el segundo capítulo de [10] y los teoremas 2.2.3, 2.2.4 y 2.2.6 de dicho trabajo a los espacios $\mathbb{R}^n$ y $\mathbb{R}^m$, se logra concluir que la transformación lineal de Moore-Penrose de $T : \mathbb{R}^n \to \mathbb{R}^m$ es precisamente el operador $\tilde{T}$ del resultado anterior; de este modo es claro que $[T^{\dagger}] = [T]^{\dagger}$. (No se realiza la prueba formal por la extensión que ella daría a este trabajo).*

Capítulo 3

Aplicaciones

Con los resultados clásicos que se han visto en el primer capitulo, se puede apreciar que cuando el problema

$$\left\{\begin{array}{l} \quad A\,\overline{x} = \overline{b} \\ \text{S.a:} \\ \quad A \in \mathcal{M}_{m\times n}(\mathbb{R}) \\ \quad \overline{b} \in \mathbb{R}^m \\ \quad \overline{x} \in \mathbb{R}^n \text{ (Vector incógnita)} \end{array}\right. \tag{3.1}$$

es pequeño, es posible determinar si tiene o no una solución manipulando las ecuaciones que lo conforman, y en caso de tenerlas a través de dicha manipulación pueden ser determinadas. Sin embargo para poder resolver sistemas "grandes.[es] necesario disponer de algún criterio claro y sistemático que de alguna u otra forma pueda ser programado en un ordenador, de modo que el problema anterior pueda ser resuelto en forma rápida y segura computacionalmente hablando.

En este capítulo no se pretende obtener un algoritmo específico pero si solidificar la base teórica de un buen criterio de aproximación del análisis numérico llamado mínimos cuadrados y probar que tan fuerte es la Inversa de Moore-Penrose construida en el segundo capitulo, a la hora de resolver sistemas de ecuaciones y cual es su aplicación sobre mínimos cuadrados.

3.1. Sistema de ecuaciones y mínimos cuadrados

Por definición, se sabe que el problema (3.1) es compatible, siempre y cuando $C.S(A, \overline{b}) \neq \emptyset$, de este modo se tiene que:

$$\exists \overline{x}_0 \in C.S(A, \overline{b}) \;:\Rightarrow\; A\overline{x}_0 = \overline{b} \;\Rightarrow\; A\overline{x}_0 - \overline{b} = \overline{0}$$

Así, si se considera que:

$$\overline{e} := \overline{b} - A\overline{x}_0 \;\Rightarrow\; \overline{e} = \overline{0} \;\Rightarrow\; ||\overline{e}|| := ||\overline{b} - A\overline{x}_0|| = 0 \tag{3.2}$$

Ahora bien, si el problema (3.1) fuera incompatible y se deseara escoger un $\tilde{x}$ en $\mathbb{R}^n$ que de alguna manera pueda ser considerado como "una solución" del mismo, por lo visto en (3.2) dicho candidato a "solución" debería ser aquel que haga que la norma del vector $\overline{e} := \overline{b} - A\tilde{x}$ se encuentre tan cerca de cero como sea posible, es decir aquel que minimice el vector $||\overline{e}|| := ||\overline{b} - A\tilde{x}||$.

De hecho, una vez que $\tilde{x}$ haya sido calculado, se puede determinar el valor de $||\overline{b} - A\tilde{x}||$ para poder ver que tan cerca de ser compatile se encuentra el sistema, obviamente cuanto más cerca de cero se encuentre $||\overline{b} - A\tilde{x}||$, más cercano a ser compatible se encuentra el problema (3.1).

Lo antes mencionado, motiva la siguiente definición.

Solución por mínimos cuadrados de un sistema lineal

Definición 3.1. *Sea $A_{m\times n}(\mathbb{R})$ y $\overline{b}$ un vector de $\mathbb{R}^m$. El vector $\tilde{x} \in \mathbb{R}^n$ es una solución por minamos cuadrados del sistema $A\overline{x} = \overline{b}$, si y sólo si:*

$$\forall\, \overline{x} \in \mathbb{R}^n : ||\overline{b} - A\tilde{x}|| \leq ||\overline{b} - A\overline{x}||$$

Nota 3.1. *El vector $\overline{e} := \overline{b} - A\tilde{x}$ habitualmente es llamado vector error de mínimos cuadrados y su norma es llamada error de mínimos cuadrados.*

Teorema 3.1.
Sea S un subespacio de dimensión finita de $\mathbb{R}^n$. Si $\overline{u}_0$ es un vector de $\mathbb{R}^n$, entonces:

$$\exists!\, \overline{p} \in S / ||\overline{u}_0 - \overline{p}|| \leq ||\overline{u}_0 - \overline{s}||, \forall \overline{s} \in S$$

Demostración:

Como S es un subespacio finito dimensional de $\mathbb{R}^n$, el teorema 1.5 garantiza que:

$$\exists!\, \overline{p} \in S / \overline{p} = Proy_S(\overline{u}_0)$$

Luego como $\overline{p} \in S$ es la proyección ortogonal de $\overline{u}_0$ sobre S, entonces por el teorema 1.6 se tiene que:

$$\forall\, \overline{v} \in S : ||\overline{u}_0 - \overline{p}|| \leq ||\overline{u}_0 - \overline{v}|| \tag{3.3}$$

Ahora bien, sea $\overline{p}_1 \in S$ otro vector distinto a $\overline{p}$ un vector tal que:

$$\forall \overline{w} \in S : ||\overline{u}_0 - \overline{p}_1|| \leq ||\overline{u}_0 - \overline{w}|| \tag{3.4}$$

Al tomar en forma particular $\overline{v} = \overline{p}_1$ y $\overline{w} = \overline{p}$, de (3.3) y (3.4), se obtiene que:

$$||\overline{u}_0 - \overline{p}|| \leq ||\overline{u}_0 - \overline{p}_1|| \leq ||\overline{u}_0 - \overline{p}|| \;\Rightarrow\; ||\overline{u}_0 - \overline{p}_1|| = ||\overline{u}_0 - \overline{p}|| \tag{3.5}$$

Ahora, como $\overline{p} = Proy_S(\overline{u}_0)$ por definición se tiene que:

$$\overline{u}_0 - \overline{p} \perp S \;:\Leftrightarrow\; \overline{u}_0 - \overline{p} \perp \overline{s}, \forall \overline{s} \in S$$

Luego, como $\overline{p}$ y $\overline{p}_1$ son elementos de S, según lo anterior se tendrá que:

$$\begin{aligned}
&\overline{u}_0 - \overline{p} \perp \overline{p} - \overline{p}_1\\
\overset{\text{th.1.4}}{\Rightarrow}\; &||(\overline{u}_0 - \overline{p}) + (\overline{p} - \overline{p}_1)||^2 = ||\overline{u}_0 - \overline{p}||^2 + ||\overline{p} - \overline{p}_1||^2\\
\Rightarrow\; &||\overline{u}_0 - \overline{p}_1||^2 = ||\overline{u}_0 - \overline{p}||^2 + ||\overline{p} - \overline{p}_1||^2\\
\overset{(3.5)}{\Rightarrow}\; &||\overline{u}_0 - \overline{p}_1||^2 = ||\overline{u}_0 - \overline{p}_1||^2 + ||\overline{p} - \overline{p}_1||^2\\
\Rightarrow\; &||\overline{p} - \overline{p}_1||^2 = 0\\
\Rightarrow\; &\overline{p} = \overline{p}_1 \quad (\Rightarrow\Leftarrow)
\end{aligned}$$

Lo anterior es una contradicción porque se asumió que $\overline{p} \neq \overline{p}_1$, por tanto es claro que:

$$\exists!\, \overline{p} \in S / ||\overline{u}_0 - \overline{p}|| \leq ||\overline{u}_0 - \overline{s}||, \forall \overline{s} \in S$$

■

Observación 3.1. *El teorema que se acaba de probar es importante ya que él está afirmando que dado un vector $\overline{u}_0 \in \mathbb{R}^n$ fijo y un subespacio finito dimensional S, existe un único vector $\overline{p} \in S$ que aproxima "mejor" a $\overline{u}_0$ en el sentido de que $||\overline{u}_0 - \overline{p}|| \leq ||\overline{u}_0 - \overline{s}||, \forall \overline{s} \in S$. Además como ya se intuía, dicho $\overline{p}$ justamente es la proyección ortogonal de $\overline{u}_0$ sobre S.*

Teorema 3.2.
Sea $A_{m\times n}(\mathbb{R})$, $\overline{b}_0$ un vector fijo de $\mathbb{R}^m$ y $\widehat{b} := Proy_W(\overline{b}_0)$ donde $W := Col(A)$.

El vector $\widehat{x}_0 \in \mathbb{R}^n$ es solución de mínimos cuadrados del sistema $A\overline{x} = \overline{b}_0$ si y sólo si $A\widehat{x}_0 = \widehat{b}$.

Demostración:

$(\Rightarrow)$

Por hipótesis se tiene que:

$$
\begin{aligned}
\widehat{b} &:= Proy_{_W}(\overline{b}_0)\,,\ W := Col(A)\\
&\overset{\text{th.1.6}}{\Rightarrow} ||\overline{b}_0 - \widehat{b}|| \leq ||\overline{b}_0 - \overline{w}||, \forall \overline{w} \in W := Col(A)
\end{aligned}
\tag{3.6}
$$

Ahora bien, como:

$$
\begin{aligned}
\widehat{b} := Proy_{_W}(\overline{b}_0) \ &:\Rightarrow\ \widehat{b} \in W\\
&:\Leftrightarrow\ \widehat{b} \in Col(A)\\
&\overset{\text{th.1.18}}{\Rightarrow}\ \exists\, \overline{x}_0 \in \mathbb{R}^n / A\overline{x}_0 = \widehat{b}
\end{aligned}
\tag{3.7}
$$

Por otro lado, si $\widehat{x}_0 \in \mathbb{R}^n$ es solución de mínimos cuadrados del sistema $A\overline{x} = \overline{b}_0$, entonces por definición se tiene que:

$$
||\overline{b}_0 - A\widehat{x}_0|| \leq ||\overline{b}_0 - A\overline{x}||\,,\ \forall\, \overline{x} \in \mathbb{R}^n \tag{3.8}
$$

Ahora, como para cualquier $\tilde{x} \in \mathbb{R}^n$ se verifica que $A\tilde{x} \in Col(A)$ entonces en (3.6) se tiene que:

$$
||\overline{b}_0 - \widehat{b}|| \leq ||\overline{b}_0 - A\widehat{x}_0|| \tag{3.9}
$$

Asimismo, como (3.8) se verifica para cualquier $\overline{x} \in \mathbb{R}^n$, en particular debe verificarse para $\overline{x}_0$, obteniendo así lo siguiente.

$$
\begin{aligned}
&||\overline{b}_0 - A\widehat{x}_0|| \leq ||\overline{b}_0 - A\overline{x}_0|| \overset{(3.7)}{\leq} ||\overline{b}_0 - \widehat{b}||\\
\Rightarrow\ &||\overline{b}_0 - A\widehat{x}_0|| \leq ||\overline{b}_0 - \widehat{b}||\\
\overset{(3.9)}{\Rightarrow}\ &||\overline{b}_0 - A\widehat{x}_0|| \leq ||\overline{b}_0 - \widehat{b}|| \leq ||\overline{b}_0 - A\widehat{x}_0||\\
\Rightarrow\ &||\overline{b}_0 - A\widehat{x}_0|| = ||\overline{b}_0 - \widehat{b}|| \overset{(3.6)}{\leq} ||\overline{b}_0 - \overline{w}||, \forall \overline{w} \in W := Col(A)\\
\Rightarrow\ &||\overline{b}_0 - A\widehat{x}_0|| \leq ||\overline{b}_0 - \overline{w}||, \forall \overline{w} \in W := Col(A)\\
\Rightarrow\ &\exists \overline{q}_0 := A\widehat{x}_0 \in Col(A)/||\overline{b}_0 - \overline{q}_0|| \leq ||\overline{b}_0 - \overline{w}||, \forall \overline{w} \in W := Col(A)
\end{aligned}
\tag{3.10}
$$

Ahora, el teorema 3.1 garantiza que existe un único elemento en $W := Col(A)$ que verifica (3.10) y es el vector proyección, es decir:

$$
\begin{aligned}
&\overline{q}_0 = Proy_{_W}(\overline{b}_0)\,,\ W := Col(A)\\
\Rightarrow\ &A\widehat{x}_0 = Proy_{_W}(\overline{b}_0)\,,\ W := Col(A)\\
\Rightarrow\ &A\widehat{x}_0 = \widehat{b}\,,\ \widehat{b} = Proy_{_W}(\overline{b}_0)
\end{aligned}
$$

($\Leftarrow$)

Por hipótesis se tiene que:

$$A\widehat{x}_0 = \widehat{b}\ ,\ \widehat{b} = Proy_W(\bar{b}_0)\ \ W := Col(A)$$

Ahora, como:

$$\begin{aligned} \widehat{b} := Proy_W(\bar{b}_0) &\overset{\text{th.1.6}}{\Rightarrow} ||\bar{b}_0 - \widehat{b}|| \leq ||\bar{b}_0 - \overline{w}||, \forall \overline{w} \in W := Col(A) \\ &\Leftrightarrow ||\bar{b}_0 - A\widehat{x}_0|| \leq ||\bar{b}_0 - \overline{w}||, \forall \overline{w} \in W := Col(A) \end{aligned} \tag{3.11}$$

Luego, como (3.11) se verifica para cualquier $\overline{w} \in W := Col(A)$, en forma particular también se verificará para cualquier $\overline{x} \in \mathbb{R}^n$, ya que dado $\overline{x} \in \mathbb{R}^n$ el teorema 1.2 garantiza que $A\overline{x} \in Col(A)$. Por lo tanto:

$$||\bar{b}_0 - A\widehat{x}_0|| \leq ||\bar{b}_0 - A\overline{x}||, \forall \overline{x} \in \mathbb{R}^n$$

■

Teorema 3.3.

Sea $A_{m\times n}(\mathbb{R})$ y $\bar{b} \in \mathbb{R}^m$. Los siguientes enunciados son equivalentes.

1. *Existe $\tilde{x}_0 \in \mathbb{R}^n$ de modo que, para cualquier $\overline{x} \in \mathbb{R}^n$ se tiene que $||\bar{b} - A\tilde{x}_0|| \leq ||\bar{b} - A\overline{x}||$.*
2. *Existe $\tilde{y}_0 \in Col(A)$ de modo que para cualquier $y \in Col(A)$ se verifica que $||\bar{b} - \tilde{y}_0|| \leq ||\bar{b} - y||$.*

Demostración:

Afirmación 1. *Si $\bar{b} \in \mathbb{R}^m$ es un vector tal que $\bar{b} \in Col(A)$ entonces existe $\bar{\xi}_0 \in R^n / A\bar{\xi}_0 = \bar{b}$*

En efecto se verifica, por lo señalado en el teorema 1.2

Afirmación 2. $\forall \overline{x} \in \mathbb{R}^n : A\overline{x} \in Col(A)$

En efecto se verifica, por lo señalado en el teorema 1.18

(1) $\Rightarrow$ (2)

Como hipótesis se tiene que:

$$\exists \tilde{x}_0 \in R^n / ||\bar{b} - A\tilde{x}_0|| \leq ||\bar{b} - A\overline{x}||, \forall \overline{x} \in \mathbb{R}^n \tag{3.12}$$

Notar que, como:

$$\exists \tilde{x}_0 \in \mathbb{R}^n \overset{\text{Af.2}}{\Rightarrow} A\tilde{x}_0 \in Col(A) :\Rightarrow\ \tilde{y}_0 \in Col(A) / \tilde{y}_0 = A\tilde{x}_0 \tag{3.13}$$

Por tanto de (3.12) se sigue que:

$$\exists \tilde{y}_0 \in Col(A)/||\overline{b} - \tilde{y}_0|| \leq ||\overline{b} - A\overline{x}||, \forall \overline{x} \in \mathbb{R}^n \tag{3.14}$$

Por otro lado:

$$\forall y \in \mathbb{R}^m: \; y \in Col(A) \overset{\text{Af.1}}{\Rightarrow} \exists \overline{\xi}_0 \in \mathbb{R}^n / y = A\overline{\xi}_0$$

Por consiguiente, como (3.14) se cumple para todo $\overline{x} \in \mathbb{R}^n$ en forma particular se cumplirá para $\overline{\xi}_0$, de este modo se obtiene que:

$$\exists\, \tilde{y}_0 \in Col(A)/||b - \tilde{y}_0|| \leq ||b - y||, \forall y \in Col(A)$$

(2) ⇒ (1)

Se tiene como hipótesis que:

$$\exists \tilde{y}_0 \in Col(A)/||b - \tilde{y}_0|| \leq ||b - \overline{y}||, \forall \overline{y} \in Col(A) \tag{3.15}$$

Ahora, por la afirmación 2 se tiene que:

$$\begin{aligned} \forall \overline{x} : \overline{x} \in \mathbb{R}^n \;&\Rightarrow\; A\overline{x} \in Col(A) \\ &\overset{(3.15)}{\Rightarrow} \exists \tilde{y}_0 \in Col(A)/||b - \tilde{y}_0|| \leq ||b - A\overline{x}||, \forall \overline{x} \in \mathbb{R}^n \end{aligned} \tag{3.16}$$

Por otro lado por la afirmación 1 se sabe que, como:

$$\begin{aligned} \tilde{y}_0 \in Col(A) \;&\Rightarrow\; \exists \tilde{x}_0 \in \mathbb{R}^n / \tilde{y}_0 = A\tilde{x}_0 \\ &\overset{(3.16)}{\Rightarrow} \exists \tilde{x}_0 \in \mathbb{R}^n / ||b - A\tilde{x}_0|| \leq ||b - A\overline{x}||, \forall \overline{x} \in \mathbb{R}^n \end{aligned}$$

■

Teorema 3.4. *Teorema de mínimos cuadrados*
Si $A_{m\times n}(\mathbb{R})$ y $\overline{b} \in \mathbb{R}^m$ entonces se verifica lo siguiente.

1. *El sistema $A\overline{x} = \overline{b}$ siempre tiene al menos una solución de mínimos cuadrados.*

2. *Dado el vector $\tilde{x}_0 \in \mathbb{R}^n$ se tiene que:*
   $\tilde{x}_0$ es una solución de mínimos cuadrados del sistema $A\overline{x} = \overline{b}$, si y sólo si $\tilde{x}_0$ es una solución del sistema $A^TA\overline{x} = A^T\overline{b}$ (llamado sistema de ecuaciones normales.)

3. *Las columnas de A, son vectores linealmente independientes si y sólo si A^TA es una matriz invertible.*
   En este caso, la solución de mínimos cuadrados del sistema $A\overline{x} = \overline{b}$ es única y está dada por $\tilde{x}_0 = (A^TA)^{-1}A^T\overline{b}$.

Demostración:

(1)

Dada la matriz $A_{m\times n}(\mathbb{R})$ por definición se sabe que $Col(A)\,\tilde{\subset}\,\mathbb{R}^m$, por tanto se tiene que:

$$\begin{aligned}
&\overline{b}\in\mathbb{R}^m \ \wedge \ \ Col(A)\ \tilde{\subset}\ \mathbb{R}^m \\
&\overset{\text{th.3.1}}{\Rightarrow} \exists!\overline{p}\in Col(A)/||\overline{b}-\overline{p}||\leq||\overline{b}-\overline{w}||,\forall\overline{w}\in Col(A) \\
&\overset{\text{th.3.3}}{\Rightarrow} \exists\tilde{x}_0\in\mathbb{R}^n/||\overline{b}-A\tilde{x}_0||\leq||\overline{b}-A\overline{x}||,\forall\overline{x}\in\mathbb{R}^n
\end{aligned}$$

Esto quiere decir que existe la solución de mínimos cuadrados.

(2)

$(\Rightarrow)$

Sea $A\overline{x}=\overline{b}$ un sistema lineal, si $\tilde{x}_0\in\mathbb{R}^n$ es una solución de mínimos cuadrados entonces por el teorema 3.2 se tiene que:

$$A\tilde{x}_0=\widehat{b}\ ,\ \widehat{b}=Proy_{W}(\overline{b})\ \ W:=Col(A) \tag{3.17}$$

Ahora, como:

$$\widehat{b}=Proy_{W}(\overline{b})\ \ :\Rightarrow\ \ \overline{b}-\widehat{b}\perp W:=Col(A) \tag{3.18}$$

Notar ahora que (3.18) está diciendo que el vector $\overline{b}-\widehat{b}\in\mathbb{R}^m$ es ortogonal a cualquier columna de la matriz A, por lo tanto, si:

$$\begin{aligned}
&A=\left[\ Col_1(A)\ \ \vdots\ \ Col_2(A)\ \ \vdots\ \ \cdots\ \ \vdots\ \ Col_n(A)\ \right] \\
&\Rightarrow\forall j\in\overline{\{1,n\}}:\left\langle Col_j(A),\overline{b}-\widehat{b}\right\rangle=0 \\
&\Rightarrow\forall j\in\overline{\{1,n\}}:[Col_j(A)]^T\ (\overline{b}-\widehat{b})=0 \qquad (3.19)
\end{aligned}$$

Luego, como:

$$A = \left[\begin{array}{ccccccc} Col_1(A) & \vdots & Col_2(A) & \vdots & \cdots & \vdots & Col_n(A) \end{array}\right]$$

$$\overset{\text{th.}1.45}{\Rightarrow} A^T = \left[\begin{array}{c} (Col_1(A))^T \\ \cdots\cdots\cdots\cdots \\ (Col_2(A))^T \\ \cdots\cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots\cdots \\ (Col_n(A))^T \end{array}\right]$$

$$\Rightarrow A^T(\overline{b} - \widehat{b}) = \left[\begin{array}{c} (Col_1(A))^T \\ \cdots\cdots\cdots\cdots \\ (Col_2(A))^T \\ \cdots\cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots\cdots \\ (Col_n(A))^T \end{array}\right] (\overline{b} - \widehat{b})$$

$$= \left[\begin{array}{c} (Col_1(A))^T(\overline{b} - \widehat{b}) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (Col_2(A))^T(\overline{b} - \widehat{b}) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (Col_n(A))^T(\overline{b} - \widehat{b}) \end{array}\right]$$

$$\overset{(3.19)}{=} \left[\begin{array}{c} 0 \\ \cdots \\ 0 \\ \cdots \\ \vdots \\ \cdots \\ 0 \end{array}\right]_{n\times 1}$$

$$:\Rightarrow A^T(\overline{b} - \widehat{b}) = \Theta$$

$$\Rightarrow A^T\overline{b} = A^T\widehat{b} \overset{(3.17)}{=} A^T A\tilde{x}_0$$

Por lo tanto $\tilde{x}_0$ es una solución del sistema $A^T A\overline{x} = A^T\overline{b}$.

Específicamente se ha probado que, si $\tilde{x}_0$ es una solución de mínimos cuadrados del sistema $A\overline{x} = \overline{b}$, entonces $A^T A\tilde{x}_0 = A^T\overline{b}$.

($\Leftarrow$)

Se debe probar que si $\tilde{x}_0 \in \mathbb{R}^n$ es una solución del sistema $A^T A\overline{x} = A^T\overline{b}$, entonces $\tilde{x}_0$ es una solución de mínimos cuadrados del sistema $A\overline{x} = \overline{b}$.

Naturalmente, si $\tilde{x}_0$ es una solución del sistema $A^T A\overline{x} = A^T\overline{b}$, entonces:

$$\begin{aligned} A^T A\tilde{x}_0 = A^T\overline{b} &\Rightarrow A^T A\tilde{x}_0 - A^T\overline{b} = \Theta \\ &\Rightarrow A^T(A\tilde{x}_0 - \overline{b}) = \Theta \\ &\Rightarrow A^T(\overline{b} - A\tilde{x}_0) = \Theta \end{aligned} \tag{3.20}$$

Por otro lado, dado que:

$$\begin{aligned} & A \in M_{m\times n}(R) \ \wedge \ \tilde{x}_0 \in \mathbb{R}^n \\ \overset{\text{th.1.2}}{\Rightarrow}\ & A\tilde{x}_0 \in Col(A) \tilde{\subset} \mathbb{R}^m \\ \Rightarrow\ & \exists \tilde{b}_0 \in \mathbb{R}^n / \tilde{b}_0 = A\tilde{x}_0 \\ \overset{(3.20)}{\Rightarrow}\ & A^T(\overline{b} - \tilde{b}_0) = \Theta \\ \Rightarrow\ & \begin{bmatrix} (Col_1(A))^T \\ \cdots\cdots\cdots \\ (Col_2(A))^T \\ \cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots \\ (Col_n(A))^T \end{bmatrix} (\overline{b} - \tilde{b}_0) = \Theta \\ \Rightarrow\ & \begin{bmatrix} (Col_1(A))^T(\overline{b} - \tilde{b}_0) \\ \cdots\cdots\cdots\cdots\cdots \\ (Col_2(A))^T(\overline{b} - \tilde{b}_0) \\ \cdots\cdots\cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots\cdots\cdots \\ (Col_n(A))^T(\overline{b} - \tilde{b}_0) \end{bmatrix} = \Theta \\ \Rightarrow\ & \forall j \in \overline{\{1, n\}} : (Col_j(A))^T(\overline{b} - \tilde{b}_0) = 0 \\ \Rightarrow\ & \forall j \in \overline{\{1, n\}} : \left\langle Col_j(A), \overline{b} - \tilde{b}_0 \right\rangle = 0 \end{aligned}$$

Lo anterior está diciendo que $\bar{b} - \tilde{b}_0$ es un vector ortogonal a todos los vectores columna de la matriz A, por consiguiente $\bar{b} - \tilde{b}_0$ debe ser ortogonal al espacio columna de A, es decir:

$$\bar{b} - \tilde{b}_0 \perp Col(A) \ :\Rightarrow \ \bar{b} - \tilde{b}_0 \in Col(A)^{\perp}$$

Por último, como $Col(A) \tilde{\subset} \mathbb{R}^m$ entonces por el teorema 1.7 se tiene que:

$$(\forall \overline{y} \in \mathbb{R}^n) \left(\exists ! (\widehat{y}, z) \in Col(A) \times Col(A)^{\perp} \right) / \overline{y} = \widehat{y} + z \ \wedge \ \widehat{y} = Proy_{W}(\overline{y})$$

Así, dado que $\bar{b} \in \mathbb{R}^n$, se tiene entonces que:

$$\exists ! (\widehat{y}, z) \in Col(A) \times Col(A)^{\perp} / \overline{y} = \widehat{y} + z \ \wedge \ \widehat{y} = Proy_{Col(A)}(\overline{y}) \tag{3.21}$$

Por otro lado, como:

$$\bar{b} = \underbrace{\tilde{b}_0}_{\in Col(A)} + \underbrace{(\bar{b} - \tilde{b}_0)}_{\in Col(A)^{\perp}} \overset{(3.21)}{\Rightarrow} \tilde{b}_0 = Proy_{Col(A)}(\bar{b})$$

Concluyendo por tanto que:

$$A\tilde{x}_0 = \tilde{b}_0 \ \wedge \ \tilde{b}_0 = Proy_{Col(A)}(\bar{b})$$

Entonces el teorema 3.2 garantiza que $\tilde{x}_0$ es una solución de mínimos ciadrados del sistema $A\overline{x} = \bar{b}$.

(3)

Dada una matriz $A_{m\times n}(\mathbb{R})$ por definición se sabe que $rg(A) = R_f(A) = R_c(A)$ donde:

$$R_f(A) := dim[Fil(A)] \ \wedge \ R_c(A) := dim[Col(A)]$$

Así:

$$\begin{aligned} rg(A) = n \ &:\Leftrightarrow \ R_c(A) = n \\ &:\Leftrightarrow \ dim[Col(A)] = n \\ &\Leftrightarrow \ \{Col_1(A), Col_2(A), \ldots, Col_n(A)\} \ son \ L.I. \end{aligned} \tag{3.22}$$

Por otro lado, por el teorema 1.20 se tiene que:

$$\exists (A^T A)^{-1} \in \mathcal{M}_{n\times n}(\mathbb{R}) \ \Leftrightarrow \ rg(A^T A) = n \overset{\text{th.2.12}}{\Leftrightarrow} rg(A) = rg(A^T A) = n \tag{3.23}$$

Así, de (3.22) y (3.23) se concluye que:

$$\begin{aligned}\exists (A^T A)^{-1} \in \mathcal{M}_{n\times n}(\mathbb{R}) &\Leftrightarrow \{Col_1(A), Col_2(A), \ldots, Col_n(A)\} \ \textit{son L.I.}\\ &\Leftrightarrow \textit{Las } n \textit{ columnas de } A \textit{ son L.I}\end{aligned}$$

Como consecuencia de esto, se tiene que, si $\tilde{x}_0$ es una solución de mínimos cuadrados del sistema $A\overline{x} = \overline{b}$, por lo probado en *(2)* se verifica que $A^T A\tilde{x}_0 = A^T\overline{b}$, donde si existe $(A^T A)^{-1}$, por teorema se tendrá que $\tilde{x}_0 = (A^T A)^{-1}A^T\overline{b}$.

■

Teorema 3.5.
Sea $A_{m\times n}(\mathbb{R})$ y $\overline{b} \in \mathbb{R}^m$ un vector fijo de $\mathbb{R}^m$. Si $A\overline{x} = \overline{b}$ es un sistema de ecuaciones lineales donde $\overline{x} \in \mathbb{R}^n$, entonces se tiene que:

1. *Si $m = n$ y el sistema $A\overline{x} = \overline{b}$ tiene solución única entonces $\overline{u}_0 := A^\dagger\overline{b}$ es la solución del sistema.*
2. *Si $A\overline{x} = \overline{b}$ es un sistema compatible indeterminado, entonces $\overline{u}_0 := A^\dagger\overline{b}$ es la solución particular del sistema.*
3. *Si $A\overline{x} = \overline{b}$ es un sistema incompatible, entonces $\overline{u}_0 := A^\dagger\overline{b}$ es una solución de mínimos cuadrados.*
4. *Si $A\overline{x} = \overline{b}$ es un sistema incompatible y existe una única solucion de mínimos cuadrados, entonces $\overline{u}_0 := A^\dagger\overline{b}$ es dicha solución de mínimos cuadrados.*

Demostración:

(1)

Bajo estas condiciones el teorema 1.16 garantiza que la solución del sistema es única y está dada por el vector:

$$\overline{x}_0 = A^{-1}\overline{b} \overset{\text{th.2.24}}{\Rightarrow} \overline{x}_0 = A^\dagger\overline{b} := \overline{u}_0$$

Esto implica que el vector $\overline{u}_0 := A^\dagger\overline{b}$ es solución del sistema $A\overline{x} = \overline{b}$ y como la solución es única entonces $\overline{u}_0$ es dicha solución.

(2)

Bajo estas condiciones por definición se tiene que:

$$\begin{aligned}&\exists \overline{x}_0 \in \mathbb{R}^n / A\overline{x}_0 = \overline{b}\\ \Rightarrow\ &\overline{b} = A\overline{x}_0 := AA^\dagger \underline{A\overline{x}_0} = A \underline{A^\dagger\overline{b}} := A\overline{u}_0\end{aligned}$$

Lo anterior implica que $A\overline{u}_0 = \overline{b}$, es decir $\overline{u}_0 := A^{\dagger}\overline{b}$ es una solución particular del sistema.

(3)

Dado que el sistema $A\overline{x} = \overline{b}$ es incompatible, por el teorema de mínimos cuadrados se sabe que existe $\tilde{x}_0 \in \mathbb{R}^n$ que es solución de mínimos cuadrados de dicho sistema. Además el mismo resultado garantiza que $\tilde{x}_0$ es solución de mínimos cuadrados si y sólo si $A^T A\tilde{x}_0 = A^T\overline{b}$.

Ahora bien, como:

$$\begin{aligned}\overline{u}_0 := A^{\dagger}\overline{b} \;&\Rightarrow\; A^T A\overline{u}_0 := A^T AA^{\dagger}\overline{b} \overset{\text{th.2.23}}{=} A^T\overline{b}\\ &\Rightarrow\; A^T A\overline{u}_0 = A^T\overline{b}\end{aligned}$$

De esto y lo antes comentado, queda claro que $\overline{u}_0 := A^{\dagger}\overline{b}$ es una solución de mínimos cuadrados.

(4)

Bajo estas condiciones el teorema de mínimos cuadrados garantiza que la única solución está dada por:

$$\tilde{x}_0 = (A^T A)^{-1}A^T\overline{b} \overset{\text{th.2.24}}{=} (A^T A)^{\dagger}A^T\overline{b} \overset{\text{th.2.23}}{=} A^{\dagger}\overline{b}$$

Por lo tanto $\overline{u}_0 := A^{\dagger}\overline{b}$ es la única solución de mínimos cuadrados.

■

Teorema 3.6.
Sea $A_{m\times n}(\mathbb{R})$ y $\overline{b} \in \mathbb{R}^m$ un vector fijo de $\mathbb{R}^m$. Si $A\overline{x} = \overline{b}$ es un sistema compatible donde $\overline{x} \in \mathbb{R}^n$ y $\overline{u}_0 := A^{\dagger}\overline{b}$ es una solución particular de dicho sistema, entonces $\overline{u}_0$ es una solución de norma mínima.

Demostración:

Como $\overline{u}_0 := A^{\dagger}\overline{b}$ es una solución particular del sistema $A\overline{x} = \overline{b}$ y además dicho sistema es compatible entonces el teorema 1.15 garantiza que:

$$Sol(A, \overline{b}) = \overline{u}_0 + Sol(A, \overline{0}) \tag{3.24}$$

Ahora bien, por teorema se sabe que:

$$\begin{aligned}
&\overline{u}_0 := A^{\dagger}\overline{b} \in Col(A^{\dagger}) \overset{th.2.31}{\Leftrightarrow} \overline{u}_0 \in Col(A^T) \overset{\text{th.}1.22}{\Leftrightarrow} \overline{u}_0 \in [Nul(A)]^{\perp} \\
&:\Rightarrow \overline{u}_0 \perp Nul(A) \\
&:\Rightarrow \overline{u}_0 \perp \xi \,,\, \forall\, \xi \in Nul(A) \\
&\overset{\text{th.}1.4}{\Rightarrow} \forall\, \xi \in Nul(A) : ||\overline{u}_0 + \xi||^2 = ||\overline{u}_0||^2 + ||\xi||^2 \geq ||\overline{u}_0||^2 \\
&\Rightarrow ||\overline{u}_0||^2 \leq ||\overline{u}_0 + \xi||^2, \forall \xi \in Nul(A) \\
&\overset{(3.24)}{\Rightarrow} ||\overline{u}_0||^2 \leq ||\overline{x}_0||^2, \forall \overline{x}_0 \in Sol(A, \overline{b}) \\
&\Rightarrow ||\overline{u}_0|| \leq ||\overline{x}_0||, \forall \overline{x}_0 \in Sol(A, \overline{b})
\end{aligned}$$

Por lo tanto $\overline{u}_0 := A^{\dagger}\overline{b}$ es una solución de norma mínima.

■

Teorema 3.7.
Sea $A_{m\times n}(\mathbb{R})$ y $\overline{b} \in \mathbb{R}^m$ un vector fijo de $\mathbb{R}^m$. Si $A\overline{x} = \overline{b}$ es un sistema incompatible entonces $\overline{u}_0 := A^{\dagger}\overline{b}$ es una solución es una solución de mínimos cuadrados de norma mínima.

Demostración:

En efecto, como:

$$\begin{aligned}
\overline{u}_0 := A^{\dagger}\overline{b} \;&\Rightarrow\; A^T A\overline{u}_0 := A^T A A^{\dagger}\overline{b} \overset{\text{th.}2.23}{=} A^T\overline{b} \\
&\Rightarrow\; A^T A\overline{u}_0 = A^T\overline{b}
\end{aligned}$$

Por tanto $\overline{u}_0 := A^{\dagger}\overline{b}$ es una solución particular del sistema $A^T A\overline{x} = A^T\overline{b}$, así, por el teorema de mínimos cuadrados se puede asegurar que $\overline{u}_0 := A^{\dagger}\overline{b}$ es una solución de mínimos cuadrados del sistema $A\overline{x} = \overline{b}$, además como es una solución particular del sistema $A^T A\overline{x} = A^T\overline{b}$, el teorema anterior garantiza que $\overline{u}_0$ es la solución de norma mínima.

Por lo tanto $\overline{u}_0 := A^{\dagger}\overline{b}$ es la solución de mínimos cuadrados de norma mínima.

■

Conclusiones

El análisis que se ha realizado sobre la inversa generalizada de Moore-Penrose de una matriz $A_{m\times n}(\mathbb{R})$ cualquiera, ha dado como resultado la existencia de la misma, a través de la descomposición en valores singulares de ella.

La unicidad de la inversa de Moore-Penrose ha sido obtenida a partir de las cuatro ecuaciones matriciales que dicha matriz satisface por definición.

Bajo ciertas condiciones ha sido posible obtener algunas expresiones de la inversa de Moore-Penrose.

Las propiedades que satisface la inversa de Moore-Penrose son una generalización de las propiedades que tiene la inversa de una matriz, cuando ella existe.

Teóricamente siempre es posible resolver un sistema de ecuaciones usando la inversa de Moore-Penrose.

La inversa de Moore-Penrose es un objeto matemático que actualmente se sigue estudiando no sólo para fundamentar el análisis numérico, puesto que dicho concepto se ha generalizado para espacios complejos, para anillos, álgebras C^*, etc.

Sugerencias

En el presente trabajo se ha formalizado la teoría respectiva a la inversa de Moore-Penrose para matrices con entradas reales, sin embargo, esta teoría puede ser generalizada al campo complejo. Se sugiere generalizar lo realizado en este trabajo al caso complejo y comparar los resultados que allí se obtienen.

Existen otros resultados respecto a la inversa de Moore-Penrose que en este trabajo no han sido colocados, debido a que la información respecto a la formalidad de esta teoría se encuentra dispersa, se sugiere unificar y complementar lo que en el presente trabajo se ha elaborado.

En los textos de análisis númerico moderno, existe un resultado que permite calcular en forma explícita la inversa de Moore-Penrose de cualquier matriz, pero que en este trabajo no se ha demostrado por la extensión que el mismo ya presenta, realizar la demostración formal de dicho resultado sería completamente interesante para los fines del mismo.

Referencias bibliográficas

[1] ALCAZAR, J. (2013). *Apuntes de Álgebra Lineal.* Alcalá de Henares: UNIVERSIDAD DE ALCALA DE HENARES.

[2] Asmar, J. (1995). *Tópicos en Teoría de Matrices.* Material no publicado.

[3] Cobos, J., Osuna, A., Robles, R. and Silva, B. (2008). *ÁLGEBRA LINEAL* [Apuntes académicos]. Recuperado de http://ma1.eii.us.es/Material/

[4] Del valle, J. (2011). *ÁLGEBRA LINEAL PARA ESTUDIANTES DE INGENIERÍA Y CIENCIAS.* México: McGRAW-HILL/INTERAMERICANA EDITORES.

[5] Díaz, A., Bargueño, V., Romera, C., Ruiz, M., and Tejero, L. (2008). *Álgebra lineal básica.* España: SANZ Y TORRES.

[6] Fioresi, R., and Morigi, M. (2015). *Introduzione all'algebra lineare.* Italia: C.E.A.

[7] Jeronimo, G., Sabia, J., and Tesauri, S. (2008). *Álgebra Lineal.* Buenos Aires: Universidad de Buenos Aires.

[8] Kolman, B., Hill, D. (2006). *ÁLGEBRA LINEAL.* México: Pearson Educación.

[9] Lay, D. (2007). *Álgebra lineal y sus aplicaciones.* México: Pearson Educación.

[10] López, O. (2017). *EXISTENCIA DEL INVERSO DE MOORE-PENROSE DE LA PERTURBACIÓN DE OPERADORES ACOTADOS EN ESPACIOS DE HILBERT* (Tesis de pregrado). Universidad Nacional de Trujillo, Perú.

[11] Monsalve, S. (2009). *Matemáticas básicas para economistas.* Bogotá: Universidad Nacional de Colombia.

[12] NOGUERA, W. (2009). *Algebra Lineal Para Estadísticos y Actuarios* [Apuntes académicos]. Libros Para Estudiantes Universitarios.

[13] PENROSE, R. (1954). *A GENERALIZED INVERSE FOR MATRICES.* Cambridge Philosophical Society, 406-413.

[14] POOLE, D. (2011). *Álgebra lineal una introducción moderna.* México: Cengage Learning Editores.

[15] RINCÓN, A. (2006). *ÁLGEBRA LINEAL.* México: Universidad Nacional Autónoma de México.

[16] ROJO, A. (1995). *Algebra II.* Buenos Aires: El Ateneo.

[17] SAITA, M. (2008). *Appunti di algebra lineare* [Apuntes académicos]. Recuperado de https://users.dimi.uniud.it/~giuseppe.lancia/psdir/.

[18] SANTOS, R. (2014). *Um Curso de Geometria Analítica e Álgebra Linear.* Belo Horizonte: Imprensa Universitária da UFMG.

[19] WANG, G. (2018). *Generalized Inverses: Theory and Computations.* New York: Springer.